

STIC-Biot ch/ChemLib

Fr m: Ramirez, Delia
Sent: Tuesday, July 09, 2002 6:06 PM
To: STIC-Biotech/ChemLib
Subject: case 09/687230

Hi,

I would like to request the following searches (Brasemann):

1. a standard search of seq id 1 and 2 in the protein databases (commercial and interference)
2. a standard search of seq id 2 151-313 in the protein databases (commercial and interference)

Thank you,

Delia M. Ramirez, Ph.D.
Patent Examiner - Art Unit 1652
USPTO
1911 S. Clark Street, Crystal Mall 1, 10D04, Mail room 10C01
Arlington, VA 22202
(703) 306-0288
delia.ramirez@uspto.gov

Point of Contact:
Barb O'Bryen
Technical Information Specialist
STIC CM1 6A05 308-4291

Point of Contact:
Barb O'Bryen
Technical Information Specialist
STIC CM1 6A05 308-4291

Searcher: BOB
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: 7-12-02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 15:50:17 ; Search time 49.31 Seconds
(without alignments)
2066.397 Million cell updates/sec

Title: US-09-687-230-2
3073
Sequence: 1 MGKKHKHKSHLYEYVE.....PGNMICLLGPSSSEKCLINK 589

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP virus:*
16: SP bacteriophage:*
17: SP archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2990.5	97.3	651	4 Q9NP11	Q9np11 homo sapien
2	2989	97.3	652	4 Q9UH59	Q9uh59 homo sapien
3	2744.5	89.3	651	11 Q88665	Q88665 mus musculu
4	1980	64.4	459	4 Q9BV48	Q9bv48 homo sapien
5	1413.5	46.0	351	4 Q96K44	Q96k44 homo sapien
6	699.5	22.8	861	5 Q9VIX2	Q9vix2 drosophila
7	684	22.3	501	4 Q9H8M2	Q9h8m2 homo sapien
8	506	16.5	636	5 Q17581	Q17581 caenorhabdi
9	413.5	13.5	174	11 Q9CT78	Q9ct78 mus musculu
10	391	12.7	233	4 Q9H7R9	Q9h7r9 homo sapien
11	274	8.9	362	4 Q9H505	Q9h505 homo sapien
12	252.5	8.2	715	4 Q9Y4Q3	Q9y4q3 homo sapien
13	252	8.2	56	4 Q9UN01	Q9un01 homo sapien
14	231.5	7.5	706	11 Q99JY4	Q99jy4 mus musculu
15	229	7.5	805	4 Q95692	Q95692 homo sapien
16	225.5	7.3	556	10 Q9SFX1	Q9sfx1 arabidopsis

17	222	7.2	652	10 Q9LM88	Q9lm88 arabidopsis
18	218	7.1	1430	5 Q9V4J4	Q9v4j4 drosophila
19	217.5	7.1	952	10 Q9SN19	Q9sn19 oryza sativ
20	208	6.8	2065	5 Q97068	Q97068 drosophila
21	199.5	6.5	1633	13 Q90941	Q90941 gallus gall
22	197	6.4	145	10 Q9FIA2	Q9fia2 arabidopsis
23	194.5	6.3	1582	4 Q9H301	Q9h301 homo sapien
24	194.5	6.3	1630	13 Q90753	Q90753 gallus gall
25	193.5	6.3	1568	13 Q90755	Q90755 gallus gall
26	192.5	6.3	1235	4 Q9H0E9	Q9h0e9 homo sapien
27	191.5	6.2	578	3 Q9Y7N0	Q9y7n0 schizosacch
28	191	6.2	729	13 Q90971	Q90971 gallus gall
29	191	6.2	733	13 Q73897	Q73897 gallus gall
30	189.5	6.2	236	11 Q9D007	Q9d007 mus musculu
31	189.5	6.2	1679	4 Q9HBD4	Q9hbd4 homo sapien
32	189	6.2	1582	4 Q9H2T3	Q9h2t3 homo sapien
33	189	6.2	1602	4 Q9H2T5	Q9h2t5 homo sapien
34	189	6.2	1634	4 Q9H314	Q9h314 homo sapien
35	188	6.1	2081	10 Q9LH98	Q9lh98 arabidopsis
36	187.5	6.1	757	4 Q15355	Q15355 homo sapien
37	187	6.1	1647	4 Q9HBD3	Q9hbd3 homo sapien
38	185	6.0	1022	11 Q63928	Q63928 mus sp. bry
39	183	6.0	638	3 Q07442	Q07442 saccharomyc
40	183	6.0	1865	11 Q60544	Q60544 mesocricetu
41	182	5.9	920	4 Q43178	Q43178 homo sapien
42	182	5.9	920	4 Q969M9	Q969m9 homo sapien
43	181.5	5.9	3080	5 Q9VR33	Q9vr33 drosophila
44	181	5.9	798	11 Q54795	Q54795 mus musculu
45	181	5.9	798	11 Q88411	Q88411 mus musculu

ALIGNMENTS

RESULT 1
ID Q9NP11 PRELIMINARY; PRT; 651 AA.
AC Q9NP11:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE BROMODOMAIN CONTAINING PROTEIN (NMG4).
GN BPA65.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=FETAL BRAIN;
RC Kziyshkowska J.G., Dobner T.G.;
RA "Cloning of a human bromodomain containing protein."
RT Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yu Y., Li G.Y.;
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ271881; CAB72445.1; -
DR EMBL; AF152604; AAF75126.1; -
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
SQ SEQUENCE 651 AA; 74138 MW; 2987947644C215E7 CRC64;

Query Match 97.3%; Score 2990.5; DB 4; Length 651;
Best local Similarity 99.3%; Pred. No. 8.5e-189;
Matches 576; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
1 MGKKHKHKSHLYEYVEPKLVKYGNGEYTELSTGSSGHDSLSFEDKRDHKHKND 60
|||||

```

Db 1 MGKHKHKKSDKHLVEEYERPLKLVKVGNEVELSTGSSGHDLSFEDKNDHDKKD 60
QY 61 RRRKKRKGEOIPGEERGRKRRRVKEDKKRDRVNEAEKDLQCHAPVRLDLPPEKP 120
Db 61 RRRKKRKGEOIPGEERGRKRRRVKEDKKRDRVNEAEKDLQCHAPVRLDLPPEKP 120
QY 121 LTSSLAQEEVEQPTLOEALNQLMRQLOKRDPSAFSPVPYDFIAPGYSMTIKHPMFST 180
Db 121 LTSSLAQEEVEQPTLOEALNQLMRQLOKRDPSAFSPVPYDFIAPGYSMTIKHPMFST 180
QY 181 MKERIKNNNDYOSIELKDNFKLMCTNMIYKPPETIYKKAOKLLHSGMKILSGERTIOSL 240
Db 181 MKERIKNNNDYOSIELKDNFKLMCTNMIYKPPETIYKKAOKLLHSGMKILSGERTIOSL 240
QY 241 KOSIDFMAADLOKTRKQKQDGTDSGSGEDGCWQEREDSGDAEHAHAKSPSKENKKDKD 300
Db 241 KOSIDFMAADLOKTRKQKQDGTDSGSGEDGCWQEREDSGDAEHAHAKSPSKENKKDKD 300
QY 301 MLEBDFKSNLLEBQEOBOLDIVKESGKLTIRLVNSQCEFERRRKPDGTTGLLHPVDPI 360
Db 301 MLEBDFKSNLLEBQEOBOLDIVKESGKLTIRLVNSQCEFERRRKPDGTTGLLHPVDPI 360
QY 361 VGERGYCLVRLGMITGRLQSGVNTLQGFKEKDKRNVTPVLYLNGPYSSAAPHYDSTFAN 420
Db 361 VGERGYCLVRLGMITGRLQSGVNTLQGFKEKDKRNVTPVLYLNGPYSSAAPHYDSTFAN 420
QY 421 ISKDDSDLIYSTYGEPSDLPDSFSIHEFLATCODYPYVADSLLDVLTKGHSRTIOEME 480
Db 421 ISKDDSDLIYSTYGEPSDLPDSFSIHEFLATCODYPYVADSLLDVLTKGHSRTIOEME 480
QY 481 MSLEPEDEGHTRLDTGKEMQITVEVPPGRLDSSDODRLIALKAVINFGVPEVPESEEA 540
Db 481 MSLEPEDEGHTRLDTGKEMQITVEVPPGRLDSSDODRLIALKAVINFGVPEVPESEEA 540
QY 541 EIFOKKIDETTRLLRELQEAONERLSTRPPNNICLLGPS 580
Db 541 EIFOKKIDETTRLLRELQEAONERLSTRPPNNICLLGPS 580

```

```

QY 61 RRRKKRKGEOIPGEERGRKRRRVKEDKKRDRVNEAEKDLQCHAPVRLDLPPEKP 120
Db 61 RRRKKRKGEOIPGEERGRKRRRVKEDKKRDRVNEAEKDLQCHAPVRLDLPPEKP 120
QY 121 LTSSLAQEEVEQPTLOEALNQLMRQLOKRDPSAFSPVPYDFIAPGYSMTIKHPMFST 180
Db 121 LTSSLAQEEVEQPTLOEALNQLMRQLOKRDPSAFSPVPYDFIAPGYSMTIKHPMFST 180
QY 181 MKERIKNNNDYOSIELKDNFKLMCTNMIYKPPETIYKKAOKLLHSGMKILSGERTIOSL 240
Db 181 MKERIKNNNDYOSIELKDNFKLMCTNMIYKPPETIYKKAOKLLHSGMKILSGERTIOSL 240
QY 241 KOSIDFMAADLOKTRKQKQDGTDSGSGEDGCWQEREDSGDAEHAHAKSPSKENKKDKD 300
Db 241 KOSIDFMAADLOKTRKQKQDGTDSGSGEDGCWQEREDSGDAEHAHAKSPSKENKKDKD 300
QY 301 MLEBDFKSNLLEBQEOBOLDIVKESGKLTIRLVNSQCEFERRRKPDGTTGLLHPVDPI 360
Db 301 MLEBDFKSNLLEBQEOBOLDIVKESGKLTIRLVNSQCEFERRRKPDGTTGLLHPVDPI 360
QY 361 VGERGYCLVRLGMITGRLQSGVNTLQGFKEKDKRNVTPVLYLNGPYSSAAPHYDSTFAN 420
Db 361 VGERGYCLVRLGMITGRLQSGVNTLQGFKEKDKRNVTPVLYLNGPYSSAAPHYDSTFAN 420
QY 421 ISKDDSDLIYSTYGEPSDLPDSFSIHEFLATCODYPYVADSLLDVLTKGHSRTIOEME 480
Db 421 ISKDDSDLIYSTYGEPSDLPDSFSIHEFLATCODYPYVADSLLDVLTKGHSRTIOEME 480
QY 481 MSLEPEDEGHTRLDTGKEMQITVEVPPGRLDSSDODRLIALKAVINFGVPEVPESEEA 540
Db 481 MSLEPEDEGHTRLDTGKEMQITVEVPPGRLDSSDODRLIALKAVINFGVPEVPESEEA 540
QY 541 EIFOKKIDETTRLLRELQEAONERLSTRPPNNICLLGPS 580
Db 541 EIFOKKIDETTRLLRELQEAONERLSTRPPNNICLLGPS 580

```

Query Match 97.3%; Score 2989; DB 4; Length 652;

Best Local Similarity 99.0%; Pred. No. 1.1e-188;

Matches 574; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MGKHKHKKSDKHLVEEYERPLKLVKVGNEVELSTGSSGHDLSFEDKNDHDKKD 60
 Db 1 MGKHKHKKSDKHLVEEYERPLKLVKVGNEVELSTGSSGHDLSFEDKNDHDKKD 60

Query Match 89.3%; Score 2744.5; DB 11; Length 651;

Best Local Similarity 88.4%; Pred. No. 1.3e-172;

Matches 518; Conservative 36; Mismatches 31; Indels 1; Gaps 1;

QY 1 MGKHKHKKSDKHLVEEYERPLKLVKVGNEVELSTGSSGHDLSFEDKNDHDKKD 60
 Db 1 MGKHKHKKSDKHLVEEYERPLKLVKVGNEVELSTGSSGHDLSFEDKNDHDKKD 60


```

Db 1 MGKKKKKSDRHFEYEEYKPLKLVLYKGVSEVTELTSGSGHSDSLFEDRSDDHKND 60
QY 61 RRRKKKKKEKQIPGEKGRKKRRRVKEDKKRRROVENEAEKDLOCHAPVRLDLPPEK 120
Db 61 RRRKKKKKEKQAPGEKGRKKRRRVKEDKKRRROVENEAEKDLOCHAPVRLDLPPEK 120
QY 121 LRSLSLAKOGEVEPTPLQELNOLMLOLRKDRPSAFSPFTDPIAGYSMIIRHMPDST 180
Db 121 LRSLSLAKOGEVEPTPLQELNOLMLOLRKDRPSAFSPFTDPIAGYSMIIRHMPDST 180
QY 181 MKERIKNNVQSIIEELKDNFKLMCTNAMITYNKPEITYYKAAKLLHSGMKILSQERQSL 240
Db 181 MKERIKNNVQSIIEELKDNFKLMCTNAMITYNKPEITYYKAAKLLHSGMKILSQERQSL 240
QY 241 KQSIDPMADLQTRKQKQDGTDSOGSDGCGWQREDESDGAEAFKSPSKENKKKDKD 300
Db 241 KQSIDPMADLQTRKQKQDGTDSOGSDGCGWQREDESDGAEAFKSPSKENKKKDKD 300
QY 301 MLEDFKSNLIEREOELDRIVKESGKLTRLVNSQCEFERRRKPDGTTTGLLHPVPT 360
Db 301 VLEDKMSSNSERHEQIEAVQESGKLTRLVNSQCEFERRRKPDGTTTGLLHPVPT 360
QY 361 VGEPGYCLVRLGMYTGRLOGSVNTLQGFKEKDKRNVTPVLYLNGPYSSYAPHYDSTFAN 420
Db 361 VGEPGYCLVRLGMYTGRLOGSVNTLQGFKEKDKRNVTPVLYLNGPYSSYAPHYDSTFAN 420
QY 421 ISKDDSLIYSTYGEDSDLPDSFSIHEFLATCQDYPYVADSLDLVLTGKGSRTLQEME 480
Db 421 ISKDDSLIYSTYGEDSDLPDSFSIHEFLATCQDYPYVADSLDLVLTGKGSRTLQEME 480
QY 481 MSLEDEGHTRLDQKEMEOITEVEPPGRILDSSTQDRILAKAVTNFVPEVDSSEA 540
Db 481 MSLEDEGHTRLDQKEMEOITEVEPPGRILDSSTQDRILAKAVTNFVPEVDSSEA 540
QY 541 EIFOKKIDETTRLLRELQEAONERLSTRPPNMICLLGSPSEKIL 586
Db 540 EVFORKIDETTRLLRELQEAONERLSTRPPNMICLLGSPSEKIL 586

RESULT 4
ID Q96KA4 PRELIMINARY; PRT; 459 AA.
AC Q96KA4;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE SIMILAR TO BROMODOMAIN-CONTAINING 7 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN, NEUROBLASTOMA;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC001611; AA01611.1;
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 1.
FT NON_TER 1
SQ SEQUENCE 459 AA; 51714 MW; 9EF21D1454BCE63B CRC64;

Query Match 64.4%; Score 1980; DB 4; Length 459;
Best Local Similarity 98.7%; Pred. No. 1.7e-122;
Matches 380; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 196 LKDNFKLMCTNAMITYNKPEITYYKAAKLLHSGMKILSQERQSLKQSIDPMADLQTRK 255
Db 3 IODNFKLMCTNAMITYNKPEITYYKAAKLLHSGMKILSQERQSLKQSIDPMADLQTRK 62
QY 256 QKDGTDSOGSDGCGWQREDESDGAEAFKSPSKENKKKDKDLEDFKSNLIEREO 315

```

```

Db 63 QKDGTDSOGSDGCGWQREDESDGAEAFKSPSKENKKKDKDLEDFKSNLIEREO 122
QY 316 EQLDRIVKESGKLTRLVNSQCEFERRRKPDGTTTGLLHPVPIVGEPGYCLVRLGMYT 375
Db 123 EQLDRIVKESGKLTRLVNSQCEFERRRKPDGTTTGLLHPVPIVGEPGYCLVRLGMYT 375
QY 376 GRLOGSVNTLQGFKEKDKRNVTPVLYLNGPYSSYAPHYDSTFANISKDDSLIYSTYGE 435
Db 183 GRLOGSVNTLQGFKEKDKRNVTPVLYLNGPYSSYAPHYDSTFANISKDDSLIYSTYGE 435
QY 436 DSDLPDSFSIHEFLATCQDYPYVADSLDLVLTGKGSRTLQEMENSLPDEGHTRLDPT 495
Db 243 DSDLPDSFSIHEFLATCQDYPYVADSLDLVLTGKGSRTLQEMENSLPDEGHTRLDPT 302
QY 496 GKEMEOITEVEPPGRILDSSTQDRILAKAVTNFVPEVDSSEAEIFOKKIDETTRLLR 555
Db 303 AKEMEOITEVEPPGRILDSSTQDRILAKAVTNFVPEVDSSEAEIFOKKIDETTRLLR 362
QY 556 ELQEAONERLSTRPPNMICLLGPS 580
Db 363 ELQEAONERLSTRPPNMICLLGPS 387

RESULT 5
ID Q96KA4 PRELIMINARY; PRT; 351 AA.
AC Q96KA4;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CDNA FLJ144402 FTS, CLONE HEMBA1003783, MODERATELY SIMILAR TO MUS
DE MUSCULUS BROMODOMAIN-CONTAINING PROTEIN BP75 MRNA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO, AND MAINLY HEAD;
RA Isogal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Magatsuma M., Hosolri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Iehida S., Murakawa K., Ono Y., Takiguchi S.,
RA Matanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuo Y.,
RA Niomiya K., Iwayanagi T.;
RT "NBD0 human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027308; BAB5031.1;
SQ SEQUENCE 351 AA; 39236 MW; 51E1DD2ABD7DEDA0 CRC64;

Query Match 46.0%; Score 1413.5; DB 4; Length 351;
Best Local Similarity 98.2%; Pred. No. 2.4e-85;
Matches 275; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 301 MLEDFKSNLIEREOELDRIVKESGKLTRLVNSQCEFERRRKPDGTTTGLLHPVPI 360
Db 1 MLEDFKSNLIEREOELDRIVKESGKLTRLVNSQCEFERRRKPDGTTTGLLHPVPI 60
QY 361 VGEPGYCLVRLGMYTGRLOGSVNTLQGFKEKDKRNVTPVLYLNGPYSSYAPHYDSTFAN 420
Db 61 VGEPGYCLVRLGMYTGRLOGSVNTLQGFKEKDKRNVTPVLYLNGPYSSYAPHYDSTFAN 120
QY 421 ISKDDSLIYSTYGEDSDLPDSFSIHEFLATCQDYPYVADSLDLVLTGKGSRTLQEME 480
Db 121 ISKDDSLIYSTYGEDSDLPDSFSIHEFLATCQDYPYVADSLDLVLTGKGSRTLQEME 180
QY 481 MSLEDEGHTRLDQKEMEOITEVEPPGRILDSSTQDRILAKAVTNFVPEVDSSEA 540
Db 181 MSLEDEGHTRLDQKEMEOITEVEPPGRILDSSTQDRILAKAVTNFVPEVDSSEA 239
QY 541 EIFOKKIDETTRLLRELQEAONERLSTRPPNMICLLGPS 580

```

DB 240 EIPFKKDETRRLRELOEAONERLSTRPPNMICLLGPS 279

```

RESULT 6
09VLX2 PRELIMINARY; PRT; 861 AA.
AC 09VLX2;
DT 01-MAR-2001 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG7154 PROTEIN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Phryganea; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Abril J.F., Abdayant A., An H.-J., Andrews-Plankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doop L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hosold D., Houston K.A., Howland T.J., Wei M.-H., Ibbesman C.,
RA Jallat M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasbo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington S., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Spiden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Splet E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003618; AAF52557.1;
DR FlyBase: FBgn0031947; CG7154.
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; bromodomain.1.
DR PRINTS: P00503; BROMODOMAIN.
DR SMART: SMO0297; BROMO.1.
DR PROSITE: PSS0014; BROMODOMAIN.2.1.
SQ SEQUENCE 861 AA; 95922 MW; E27D528E5F9B3AF CRC64;

```

Query Match 22.8%; Score 699.5; DB 5; Length 861;
 Best Local Similarity 27.6%; Pred. No. 7.5e-38;
 Matches 211; Conservative 127; Mismatches 219; Indels 207; Gaps 22;

```

OY 3 KKKKKKSD-KHLYEY-----VEKPLKLVLYKVGNEVTELSGSSGHD----- 45
DB 5 KKKKKKSEREREKTEESYQHODPAOLGKLLIKVGNATPEYSAKSPMDGSPPTAAE 64
OY 46 ---SLEFDKNDHDKHKKRRKKRKKKKQIPGEKGRKKRRVEDKKRRDROVENEAE 102
DB 65 AMSPVPEELQDHGHRHRHKKKKKKKK---KDRKKKHNNHE-KRHRSDHROAGS 119
OY 103 -----KDLQCH----- 108
DB 120 DEDMAGADAACGFAFPSSVAPPAADDSQDSFEMDDQSLPENILFAGITTDN 179
OY 109 -----APVRLD-----LP----- 116
DB 180 SPSPCPTKTPAPKLDLILMGSSPNSSLSQSSSLGSSPTKPLDILPSPSTGGA 239
OY 117 -----PEKPLTSSLAQK-----EVBOTPLQALNOLMROLKRPDSAF 155
DB 240 NSLNALTPKALEAPKTPSSSSSESGREPRSCVLKLOQKSPLNKLEHLRFLKRDPHQF 299
OY 156 FSPVPTDFIAPGYSMIKKHMDPSTMEKRTKNDYOSIEELKDNFKMTNMTYKPEET 215
DB 300 FAMPVTDMPAPGYSTISRPMDSTMRKIDHRYTALTEFTDFKLMCENAIKYNHVDY 359
OY 216 IYYKAANKLLHSGMKLISQERI-OSLKOSIDFMADLOKTRKODGTPDSQSGDGCWOR 274
DB 360 VYKKAARLLQYMKHLQENLRSLKPLSGYREL---TARELGELSSND-----MSR 411
OY 275 ERDSDG-----AEAHAKSPKRNKKDKMDLEKFS-----NNLREQ-DQLDR 320
DB 412 ENNDSADDEGASTGAEEPRPPAOLEEERKRTLLENAPKTHFPVYDDLTEGELLAQOV 471
OY 321 IVKESGKILTRLVNSOCEPERKRPDGTTLGLHRYVDPIVSGCYVRLGTTGRLOS 380
DB 472 AAQQAQGVNAKNAKNAKMGFLRKAKCTTILNLYEE---NGPRVATIGLVKLOK 528
OY 381 CVNTLOGFEKRNKRVLYLYNGPYSSAPHYDSTFANISKDSLDLYSTGEISDLP 440
DB 529 GSAOLQTRGVCKRNVAVTVKSLNYGAFASFAFTFDSRFSTLSAEQLVLRGTGASNAE 588
OY 441 SDPSIHEFLATQDQYVYVADSLDVLTKGHSRTL-----QEMMSLPED 487
DB 589 YASLIQFTRK-D-SNGTTTANGSLDITNGEHSKIDELYNQMLSHYEDREIEKEFE 647
OY 488 GHTRLDTGKEKEQITEVEPPGRSDSTQDRILATKAVTNFGVPEVFESEAEI----- 542
DB 648 -----ETSSQGETVAQIDQYEKYNKTHVDPKGLQSLGELGIDVSFLDGMRAEMKSYEL 701
OY 543 ---FQKKDETRRLRELOEAONERLSTRPPNMICLLGPSSEK 583
DB 702 NRRMEHLSQNLTLLEKRLVQAQHDRLS-OLPNNHGLVQPAQOE 744

```

```

RESULT 7
09H8M2 PRELIMINARY; PRT; 501 AA.
AC 09H8M2;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CDNA FLJ13441 FIS, CLONE PLACE1002775, WEAKLY SIMILAR TO PEREGRIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Ishigawa T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,

```

RESULT	9	
Q9C7T8		
ID	Q9C7T8	PRELIMINARY; PRT; 174 AA.
AC	Q9C7T8;	
DT	01-JUN-2001 (TREMBLrel, 17, Created)	
DT	01-JUN-2001 (TREMBLrel, 17, Last sequence update)	
DT	01-DEC-2001 (TREMBLrel, 19, Last annotation update)	
DE	18 DAYS EMBRYO CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,	
DE	CLONE.1190001E05, FULL INSERT SEQUENCE (FRAGMENT).	
GN	BR7.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus	
OX	NCLTaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C57BL/6J; TISSUE=EMBRYO;	
RA	MEDLINE=21085660; PubMed=11217851;	
RA	Kawai J., Shingawa A., Shidara K., Yoshino M., Itoh M., Ishii Y.,	

RA Atakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischnann W., Gaasterland T., Gissi C., King B., Kochava H.,
 RA Knehl P., Lewis S., Matsuo Y., Nikaido I., Resole G., Quackenbush J.,
 RA Schraml L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Kono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilming L.,
 RA Yushan-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.,
 RT Functional annotation of a full-length mouse cDNA collection.*;
 RL Nature 409:685-690(2001).
 DR EMBL: AK004429; BAB23299.1;
 DR MGD: MGI:1349766; Brd7.
 PT NON_TER 1 1
 SQ SEQUENCE 174 AA; 19027 MW; CF47EE3B9D719D03 CRC64;

Query Match 13.5%; Score 413.5; DB 11; Length 174;
 Best Local Similarity 75.2%; Pred. No. 6.3e-20;
 Matches 82; Conservative 11; Mismatches 15; Indels 1; Gaps 1;
 QY 478 EMMSLPDEGHTRTLDTCGKEMEQITEVEPPGRILDSSTODRLATKATNGVPEVEFDS 537
 Db 1 DLMSSPEDGCGTRALDPAKEA-ITQIEPTGRLSSQDRALQAVTTFGAPAEVFD 59
 QY 538 EEAETOKKIDETTRILRLQEAQNERLSTRPPNMCILGSPSEKCTL 586
 Db 60 EEAETOKKIDETTRILRLQEAQNERLSTRPPNMCILGSPSEKCTL 108

RESULT 10
 Q9H7R9 PRELIMINARY; PRT; 233 AA.
 AC 09H7R9;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE CDNA FLJ14330 FIS, CLONE PLACE400261, WEAKLY SIMILAR TO PERGRIN.
 DE PERGRIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN 11
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Wakamatsu A., Nakamura Y., Nagahara K., Masuo Y., Oshima A.,
 RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK024392; BAB14907.1;
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; Bromodomain; 1.
 DR PRINTS: PRO0503; BROMODOMAIN.
 DR SMART: SM00297; BROMO; 1.
 DR PROSITE: PSS0014; BROMODOMAIN.2; 1.
 SQ SEQUENCE 233 AA; 26233 MW; FBF96245BD86F4EB CRC64;

Query Match 12.7%; Score 391; DB 4; Length 233;
 Best Local Similarity 65.1%; Pred. No. 2.8e-18;
 Matches 69; Conservative 19; Mismatches 18; Indels 0; Gaps 0;

QY 130 EVEDPTIQEALINQMLRQQRKPSAFSEFPYTDFAEYSMTIKHPDSTAKEKIND 189
 Db 19 ENESPTIQDLHDHPTLRQQRKDPHGFAPYTDALAPYSMTIKHPDFTMKDKIVANE 78
 QY 190 YOSIEELKDNFKLMCTNMIYKPEYTYKRAKKLLHSGMTLSOE 235
 Db 79 YKSTEFKADPKLMCDNMATYNTNRPDYTKAKKILHAGFMMSKQ 124

RESULT 11
 Q9H505 PRELIMINARY; PRT; 362 AA.
 AC 09H505;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CDNA: FLJ23117 FIS, CLONE LNC10649.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN 11
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG;
 RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
 RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ota T.,
 RA Yamada K., Fujii Y., Ozaki K., Hideo M., Ohmori Y., Ota T., Suzuki Y.,
 RA Odayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
 RA Isogai T., Sugano S.,
 RT NEDD human cDNA sequencing project.*;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK026830; BAB15565.1;
 SQ SEQUENCE 362 AA; 39411 MW; 3B8240D7C70E44A5 CRC64;

Query Match 8.9%; Score 274; DB 4; Length 362;
 Best Local Similarity 25.7%; Pred. No. 2.5e-10;
 Matches 79; Conservative 70; Mismatches 126; Indels 32; Gaps 8;
 QY 291 SKENKKKDKMLDEKPK-----SNNLEKQQLDRIVESGKTLRLVNSQCEP 340
 Db 31 AKSKKPSSEVISCFEPBEGNACSLTDSYAEVIALVYHAADEADRLNRLPGKMKY 90
 QY 341 ERRKPDGTTTLGLHPDIVGEPGYCLVRLGTTTGRLOSGVTLQGFEDKRNKVTPL 400
 Db 91 LKRNQGSILYSVNTAEPPADDEETHPYDLSLSKLLPGFTTL-GFDERNRKYT--- 146
 QY 401 YNTGPGSSYVAHYDSTFANISKDSLDLYTYGEDSDLPSPDSHETLATQDPYVNA 460
 Db 147 FLS-SATYALSMQNNNSVFEDLKSDEWELLYSAYGDETVQCALSLIOEFVKDAGSYSKVY 205
 QY 461 DSLDLVLTGKHSRTLQENK-----MSLPDEGHTRTL-DTGKMEQITEVEPPGRILDS 513
 Db 206 DLLDLPITGDSHRTLPQKORNNYPMKPPDPAKVGDTIGDSSSYLSEMSK----- 258
 QY 514 STQDRILAKAVTNGVPEVDFSEALITFOKRLDETRRLRLQEAQNERLSTRPPGNN 573
 Db 239 STPDVSVDSLMSLSGKVKKEIDPPDSHL---NLDETTRILQDLHQAQERNGSGSPSSL 315
 QY 574 ICLIGPS 580
 Db 316 SLSMAS 322
 RESULT 12
 Q9Y4Q3 PRELIMINARY; PRT; 715 AA.
 AC 09Y4Q3;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HYPOTHETICAL 79.7 KDA PROTEIN (FRAGMENT).

GN DKEZP434B094.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 NC NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AL080149; CAB45742.1;
 DR InterPro: IPR001487; Bromodomain.
 DR InterPro: IPR000313; PWM.
 DR Pfam: PF00439; bromodomain; 1.
 DR Pfam: PF00855; PWM; 1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00297; BROMO; 1.
 DR SMART: SM00293; PWM; 1.
 DR PROSITE: PS50014; BROMODOMAIN_2; 1.
 DR KW Hypothetical protein.
 FT NON_TER
 SQ SEQUENCE 715 AA; 79702 MW; D397FF5501ED9990 CRC64;

Query Match 8.2%; Score 252.5; DB 4; Length 715;
 Best Local Similarity 26.2%; Pred. No. 1.6e-08;
 Matches 88; Conservative 53; Mismatches 102; Indels 93; Gaps 11;

QY 86 KEDKKRDRVNEAEKDLCHAPVRLDLPPEKPLTSSLAKE-----EVEQ----- 133
 DB 29 QRSOQREDEEMKAKKWKQRLRHLDERALIELLRKREKLRQVVEQVAVEL 88
 QY 134 --TPIQLALNQLMRLOLRKDPSPFSPVTDPIAGYSMTIKHPDFTMKKIKNNYQ 191
 DB 89 RLTPPLTVLRSLVDLQDLQDPARLPAQVSLKEVPDYLDHKHPDFTMKRLRLAOGVK 148
 QY 192 SIELKDNFKLCTNAMYKPEIYYKAKKLHSGMKILSQEIRISLQSIDPMADLQ 251
 DB 149 NLHFEEDFDLIDNCMKYMNADIFYRAAVRLRDGGVVLRLQAR----- 193
 QY 252 KTRKOKDGTIDISQSGEDGCWOREDSGDAEA---HAFKPSKKNKK---KDKMLIED 304
 DB 194 -----REVDSIGLEASGMHLPERPAAPRRPFSMEDVDRLD 231
 QY 305 KFKSNL---EREOROLDRI-----VKESGGKLR-RLVNSOCERERKPPDCTTLLGLH 355
 DB 232 PANRNLHLEQLRLMLMDLITCAMKSSGSKRAKLKKEIALIRNK-----LSQOH 285
 QY 356 PVDPIVGEPGYCLVRLGTTGRLOSQVNTLOGFKED 391
 DB 286 S-QPLPTGPG-----LEGFEED 301

RESULT 13
 Q9UN01 PRELIMINARY; PRT; 56 AA.
 ID Q9UN01;
 AC Q9UN01;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE FT002.
 GN FT002.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 NC NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=THYMUS;
 RA Deng Y.C., Yao L.B., Su C.Z., Lui X.P., Ji S.P., Zhang X.G., Li F.Y.,
 RA Wang Y.C., Yang M., Han J., Han Y.H., He P.;
 RT "Sequences cloned from human fetal thymus cDNA library."
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF167351; AAD46658.1;
 SQ SEQUENCE 56 AA; 6396 MW; 431E06FC3D9D7430 CRC64;

Query Match 8.2%; Score 252; DB 4; Length 56;
 Best Local Similarity 85.5%; Pred. No. 6.3e-10;
 Matches 47; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 369 VRLGTTGRLOSQVNTLOGFKEDKRNKYVPLYLYNGPYSSAPHYDSTFANISK 423
 DB 1 MRLGTTGRLOSQVNTLOGFKEDKRNKYVPLYLYNGPYSSAPHYDSTFANIKO 55

RESULT 14
 Q99JV4 PRELIMINARY; PRT; 706 AA.
 ID Q99JV4;
 AC Q99JV4;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE HYPOTHEICAL 79.9 KDA PROTEIN (FRAGMENT).
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strauberg R.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC005647; AAH05647.1;
 DR InterPro: IPR001487; Bromodomain.
 DR InterPro: IPR000313; PWM.
 DR Pfam: PF00439; bromodomain; 1.
 DR Pfam: PF00855; PWM; 1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00297; BROMO; 1.
 DR SMART: SM00293; PWM; 1.
 DR PROSITE: PS50014; BROMODOMAIN_2; 1.
 DR KW Hypothetical protein.
 FT NON_TER
 SQ SEQUENCE 706 AA; 79852 MW; 87C549B591B0639 CRC64;

Query Match 7.5%; Score 231.5; DB 11; Length 706;
 Best Local Similarity 27.5%; Pred. No. 3.7e-07;
 Matches 81; Conservative 50; Mismatches 127; Indels 37; Gaps 9;

QY 92 RDRRVENEAEKDLCHAPVRLDLPPEKPLTSSLAKE-----EVEQPLQ 137
 DB 68 RDSDDKNALKKQKLSKORLHDLERARLLVELIRKREKLRKRETIKIQOIMEMQLTPEL 127
 QY 138 EALNQLMRLOLRKDPSPFSPVTDPIAGYSMTIKHPDFTMKKIKNNYOSIELK 197
 DB 128 ILLRKTLEOLOEKDTGNIFSEVPVLDYLDHKHPDFTMKQNLDEAVRYLNPDEF 187
 QY 198 DNFKLMCTNAMYKPEIYYKAKKLHSGMKILSQEIRISLQSIDPMADLQTRKOK 257
 DB 188 EDFNLIVSNCKYNAKDTIFYRAAVRLRDGGAVLRQARQAEKMGIDFETGMHIPPULA 247
 QY 258 DGTDSQSGEDGCWOREDSGDAEAHAFKPSKKNKKDKMLIEDKFSNNERE--- 314
 DB 248 -GDEVSHHTED---VEERLVLENOKHL---PYEDOLK-----LLERDEVANSKQSVG 296
 QY 315 QEOLDRIYKESGGKLTRLVNSQCEFERRKPD--GTTTLLGRLHVPDPIVGEPGYC 367
 DB 297 RSRRAKMIKKEMTALRRRLAH-QRETRGDGPERHSGRGNLTP-----HPAAC 344

RESULT 15
 ID Q95692 PRELIMINARY; PRT; 805 AA.
 AC Q95692;
 DT 01-MAY-1999 (Tremblrel. 10, Created)

```

DN      01-MAY-1999 (TREMBLrel_10, last sequence update)
DT      01-JUN-2001 (TREMBLrel_17, last annotation update)
DE      D5J24E15.1 (PERGRIN (BR140 PROTEIN)) (FRAGMENT).
GN      D5J24E15.1.
OS      Homo sapiens (Human) .
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Palmer S.;
RL      Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL; Z84485; CAB06488.1; -.
DR      InterPro; IPR001487; Bromodomain.
DR      InterPro; IPR001965; PHD.
DR      InterPro; IPR000313; PWM.
DR      Pfam; PF00439; bromodomain; 1.
DR      Pfam; PF00628; PHD; 1.
DR      Pfam; PF00855; PWM; 1.
DR      PRINTS; PRO0503; BROMODOMAIN.
DR      SMART; SM00297; BROMO; 1.
DR      SMART; SM00249; PHD; 2.
DR      SMART; SM00293; PWM; 1.
DR      PROSITE; PS50014; BROMODOMAIN_2; 2.
FT      NON_TER      1      1
SQ      SEQUENCE      805 AA; 90851 MW; E28C0175C545334 CRC64;

```

Query Match

query Match	7.5%;	Score 229;	DB 4;	Length 805;
Best Local Similarity	26.1%;	Pred. No. 6.4e-07;		
Matches 67;	Conservative 55;	Mismatches 107;	Indels 28;	Gaps 6;

[illegible]

Search completed: July 11, 2002, 15:56:16
Job time: 359 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 15:49:34 ; Search time 55.8 Seconds
(without alignments)
324,463 Million cell updates/sec

Title: us-09-687-230-2_COPY_151_313

Perfect score: 855
Sequence: 1 DP5AFSFPVDTFAPGISM.....NKKDKDMLDEKFSNNIER 163

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	855	100.0	589	19	AAW37947 Phosphatidylinositol
2	855	100.0	667	22	AAO13504 Human polypeptide
3	855	100.0	718	22	AAU16626 Human novel secret
4	375	43.9	501	22	AA85421 Human protein sequ
5	356	41.6	405	22	AAW3712 Human polypeptide,
6	328	38.4	233	22	AA85881 Human protein sequ
7	328	38.4	280	21	AA841780 Human ORFX ORF1544
8	328	38.4	597	22	AAW58835 Human polypeptide
9	328	38.4	605	22	AAW40621 Human polypeptide
10	284.5	33.3	861	22	AB863028 Drosophila melanog
11	174	20.4	414	22	AA863767 Human prostate can

12	174	20.4	715	22	ABG12634 Novel human diagno
13	174	20.4	1058	22	AAW39231 Human polypeptide
14	174	20.4	1109	22	AAW41017 Human polypeptide
15	174	20.4	1109	22	AAW41018 Human polypeptide
16	174	20.4	1189	22	AAW39232 Human polypeptide
17	169.5	19.8	616	21	AAW28052 Arabidopsis thalia
18	169	19.8	707	22	ABG21300 Novel human diagno
19	163	19.1	1173	22	ABG21301 Novel human diagno
20	163	19.1	2543	22	ABG21295 Novel human diagno
21	160	18.7	1430	22	ABW58602 Human prostate can
22	155	18.1	270	22	AAW38879 Human prostate can
23	153	17.9	245	22	AAW38828 Novel human diagno
24	151	17.7	1924	22	ABG06008 Drosophila melanog
25	147	17.2	513	22	ABW68341 Novel human secret
26	147	17.2	969	22	AAW30116 Protein regulating
27	146	17.1	951	21	AAW58634 Drosophila melanog
28	146	17.1	2065	22	ABW6356 Novel human diagno
29	145.5	17.0	346	22	ABG06006 TATA-binding prote
30	145.5	17.0	1872	17	AAW6078 TATA-binding prote
31	145.5	17.0	1872	17	AAW25030 Novel human diagno
32	145.5	17.0	1872	18	AAW5009 TATA-binding prote
33	145.5	17.0	1886	22	ABG06009 Human TATA-binding
34	145.5	17.0	1893	15	AAW56491 Human TATA-binding
35	145.5	17.0	1893	17	AAW6082 TATA-binding prote
36	145.5	17.0	1893	18	AAW25020 Human tumour suppr
37	144	16.8	1646	21	AAW27553 Human tumour suppr
38	144	16.8	1647	21	AAW27549 Human tumour suppr
39	144	16.8	1649	21	AAW27557 Human tumour suppr
40	144	16.8	1650	21	AAW27555 Human tumour suppr
41	144	16.8	1654	22	ABW5684 Drosophila melanog
42	144	16.8	1678	21	AAW37554 Human tumour suppr
43	144	16.8	1679	21	AAW37552 Human tumour suppr
44	144	16.8	1681	21	AAW37556 Human tumour suppr
45	144	16.8	1682	21	AAW37536 Human tumour suppr

ALIGNMENTS

RESULT 1	
AAW37947	AAW37947 standard; Protein; 589 AA.
XX	XX
AC	AAW37947;
XX	XX
DT	14-SEP-1998 (first entry)
XX	XX
DE	Phosphatidylinositol-3' kinase associated protein.
XX	XX
KW	Phosphatidylinositol-3' kinase associated protein; PI3K; PIKAP;
KW	human; signal transduction; cell growth; cancer; restenosis;
XX	XX
OS	Homo sapiens.
XX	XX
FH	Key
FT	Domain
FT	Region
FT	Location/Qualifiers
FT	151..313
FT	/note="bromodomain"
FT	516..589
FT	/note="PI3K p85 binding region"
PN	W09820126-A1.
XX	XX
PD	14-MAY-1998.
XX	XX
PP	01-OCT-1997; 97WO-US15845.
XX	XX
PR	01-NOV-1996; 96US-0030103.
XX	XX
PA	(ONXX-) ONXX PHARM INC.
XX	XX
PI	Braselmann S;
XX	XX

DR WPI: 1998-286942/25.
 DR N-PSDB: AAV29267.
 XX
 PT New isolated phosphatidylinositol-3'-kinase associated protein -
 PT used to develop products for diagnosis and treatment of cell growth
 PT disorders such as restenosis or cancer
 XX
 PS Claim 10; Page 40-41; 52pp; English.

CC This polypeptide comprises human phosphatidylinositol-3' kinase
 CC (PI3K) associated protein (PIKAP), a protein that binds to the
 CC intermediate SH2 domain on the p85 regulatory subunit of PI3K, and
 CC which exhibits a bromodomain. Its amino acid sequence was deduced
 CC from a cDNA clone (see AAV29267) obtained from a HeLa library using
 CC a yeast two-hybrid assay with PI3K p85 as bait. The invention
 CC provides vectors containing nucleic acid sequences that encode
 CC PIKAP or its fragments, host cells, methods for the expression of
 CC PIKAP, and methods for using the products for the diagnosis and
 CC treatment of cell growth disorders such as restenosis or cancer.
 CC Also described is an assay for identifying agonists and antagonists
 CC of PI3K regulation. These include mutant PIKAPs that compete with
 CC native PIKAPs for binding to PI3K, antibodies, and nucleotide
 CC sequences that can be used to inhibit or enhance PIKAP gene
 CC expression. Transgenic and knock-out animals are also described.
 CC
 CC Sequence 589 AA:

Query Match 100.0%; Score 855; DB 19; Length 589;
 Best Local Similarity 100.0%; Pred. No. 1.2e-74;
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSAFSPVYDFIAPGYSMIIRKHPMDFSTMKERIKNNYOSIEELKDNFKLCTNAMY 60
 DB 151 dpaafsfpytdfiapgyssmllkhpmdfstmketkknndyqslkdnfklnctnamy 210
 QY 61 NKRETIYYKAARKLLHSGMKILSOERIQSLKSIDPMADLQKTRKQKDGTSQSGEDG 120
 DB 211 nkpetiyykaarkllhsgmkilsgeriqslksidpmadlqktrkqkdgtsgsgedg 270
 QY 121 CMQREDSGDAAHAHAFKSPSKENKKKKDKMLDCKFSNNLER 163
 DB 271 cwgqredsgdaaahafkspskenkkrkdkmlcdkfknnler 313

RESULT 2

AA013504
 ID AA013504 standard; Protein; 667 AA.

AC AA013504;

DT 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 27396.

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

PN WO200164835-A2.

PD 07-SEP-2001.

PF 26-FEB-2001; 2001WO-US04927.

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI: 2001-514838/56.
 DR N-PSDB: AAI93435.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -
 XX
 PS Claim 20; SEQ ID NO 27396; 1399pp + Sequence Listing; English.

CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 CC Sequence 667 AA:

Query Match 100.0%; Score 855; DB 22; Length 667;
 Best Local Similarity 100.0%; Pred. No. 1.4e-74;
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSAFSPVYDFIAPGYSMIIRKHPMDFSTMKERIKNNYOSIEELKDNFKLCTNAMY 60
 DB 166 dpaafsfpytdfiapgyssmllkhpmdfstmketkknndyqslkdnfklnctnamy 225
 QY 61 NKRETIYYKAARKLLHSGMKILSOERIQSLKSIDPMADLQKTRKQKDGTSQSGEDG 120
 DB 226 nkpetiyykaarkllhsgmkilsgeriqslksidpmadlqktrkqkdgtsgsgedg 285
 QY 121 CMQREDSGDAAHAHAFKSPSKENKKKKDKMLDCKFSNNLER 163
 DB 286 cwgqredsgdaaahafkspskenkkrkdkmlcdkfknnler 328

RESULT 3

AAU16626
 ID AAU16626 standard; Protein; 718 AA.

AC AAU16626;

DT 07-NOV-2001 (first entry)

DE Human novel secreted protein, Seq ID 1579.

KW Human; immunosuppressive; antiarthritic; antirheumatic;
 KW cytosolic; cardiac; vasotropic; cerebroprotective; nootropic;
 KW neuroprotective; antibacterial; virucide; fungicide; optalmalogical;
 KW vulnery; secreted protein; rheumatoid arthritis;
 KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
 KW cerebrovascular disorder; cerebral ischaemia; anglogenesis;
 KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
 KW corneal infection; wound healing; epithelial cell proliferation;
 KW skin ageing; food additive; preservative; antiproliferative.

OS Homo sapiens.

PN WO200155322-A2.

PD 02-AUG-2001.

PR 17-JAN-2001; 2001WO-US01341.

XX

XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235483.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0235802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251890.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA; Barash SC, Ruben SM;
XX
DR WPI; 2001-488783/53.
DR N-PSDB; AAS26613.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for

PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -

Claim 11; SEQ ID No 1579; 980pp; English.

The invention relates to isolated nucleic acid molecules and their encoded secreted proteins. The nucleic acids and proteins are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection, and many other disorders listed in the specification. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capability, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. The present sequence represents a novel secreted protein of the invention.

Query Match 100.0%; Score 855; DB 22; Length 718;
 Best Local Similarity 100.0%; Pred. No. 1.5e-74;
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSAFSPFVPTDIAGYSMIIKHPMDFSTMEKIKNDYOSIEELKDNFKLMCTNAMY 60
 DB 218 dpaafsfvptdiagysmiihpmdfstmekikndysieelkdnfklnctnamiy 277
 QY 61 NKRETIYYRAAKKLHSGKMTLSQERIOSLKSIDFMADLOKTRKDKGDTQSQSEDDG 120
 DB 278 nkpetiyyraakkllhsgkmtlsqerioslksidfmadlqtrkdkgdtdsqsqsdg 337
 QY 121 CMQREHEDSGDAEAFHAFKSPKSKKKDKMDLEDKFSNNLER 163
 DB 338 cmqretdsgdaeahafkspskknkdkmdledkfsnnler 380

RESULT 4
 AAB95421 standard; Protein: 501 AA.

AC AAB95421;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:17823.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-ATG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI: 2001-318749/34.

PT primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -

PS Claim 8; SEQ ID 17823; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the of an oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 501 AA;

Query Match 43.9%; Score 375; DB 22; Length 501;
 Best Local Similarity 47.5%; Pred. No. 4.6e-28;
 Matches 77; Conservative 28; Mismatches 37; Indels 20; Gaps 3;

QY 1 DSAFSPFVPTDIAGYSMIIKHPMDFSTMEKIKNDYOSIEELKDNFKLMCTNAMY 60
 DB 40 dphgffaifvptdiagysmiihpmdfstmekikndysieelkdnfklnctnamiy 99

QY 61 NKRETIYYRAAKKLHSGKMTLSQERIOSLKSIDFMADLOKTRKDKGDTQSQSEDDG 120
 DB 100 nrpdlvyyklakklhagfkmskerllakrsmfmgdm-----dfsqqaallg 149

QY 121 CMQREHEDSGDAEAFHAFKSPKSKKKDKMDLEDKFK 157
 DB 150 -----nedlaveepvpyvvetakskkpsrevlscmte 186

RESULT 5

AAM93712 standard; Protein: 405 AA.

AC AAM93712;

DT 06-NOV-2001 (first entry)

DE Human polypeptide, SEQ ID NO: 3652.

KW Human; full length cDNA; cDNA synthesis; oligo-capping.

OS Homo sapiens.

PN EP1130094-A2.

XX

PD 05-SEP-2001.
 XX 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI; 2001-524255/58.
 DR N-PSDB; AAK94661.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 PS
 PS Claim 8; SEQ ID NO 3652; 1380bp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a polypeptide
 CC encoded by a full length human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 XX
 SQ Sequence 405 AA;
 Query Match 41.6%; Score 356; DB 22; Length 405;
 Best Local Similarity 100.0%; Pred. No. 2.5e-26;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 97 MADLQKTRKQDGTSPSGSGEGCWORERSDGDAFAHAFKSPSKKKKKDKDLKEDKF 156
 DB 1 madlqktrkqgdcstsgsgsgcwgreredsgdaeahakspkknkkdkmledkf 60
 QY 157 KSNMLER 163
 DB 61 ksnmler 67
 RESULT 6
 AAB95881
 ID AAB95881 standard; Protein: 233 AA.
 AC AAB95881;
 XX
 XX 26-JUN-2001 (first entry)
 DE Human protein sequence SEQ ID NO:18979.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 OS Homo sapiens.
 OS
 XX
 XX EPI074617-A2.
 PN
 XX
 XX 07-FEB-2001.
 PD
 XX
 XX 28-JUL-2000; 2000EP-0116126.
 PF
 XX
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS
 PS Claim 8; SEQ ID 18979; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 CC
 XX
 SQ Sequence 233 AA;
 Query Match 38.4%; Score 328; DB 22; Length 233;
 Best Local Similarity 67.1%; Pred. No. 6.5e-24;
 Matches 57; Conservative 15; Mismatches 13; Indels 0; Gaps 0;
 QY 1 DPSAFSPFVTDFTAPGYSMTIKHPMDFSTMKKEIKNNDOYSIEBLKDNFKLMTNAMY 60
 DB 40 dphgffafprvdalagysmtlkhpmdfgtmkdklvaneysvtefkedfklmcdnanty 99
 QY 61 NKPEITYYKAARKLIHSGMKLISOE 85
 DB 100 nrpdtvyyklakklhagfkmmksk 124
 RESULT 7
 AAB41780
 ID AAB41780 standard; Protein: 280 AA.
 AC AAB41780;
 XX
 XX 08-FEB-2001 (first entry)
 DE Human ORFX ORF1544 polypeptide sequence SEQ ID NO:3088.
 XX
 XX
 KW Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
 KW vulnerability; antipapillary; antipapillary; antipapillary; antipapillary;
 KW anticonvulsant; osteopathic; antipapillary; immunosuppressant; cardiac;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antidiabetic;
 KW antiviral; antibacterial; antifungal; antineoplastic; antihypertensive;
 KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;

KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 XX
 PR 02-APR-1999; 99US-0127636.
 XX
 PR 05-APR-1999; 99US-0127728.
 XX
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 XX
 DR N-PSDB: AAC75989.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 XX
 PT useful for treating e.g. cancers, proliferative disorders,
 XX
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 11; Page 2306-2307; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiproliferative; antiparasitism; nootropic; neuroprotective;
 CC osteoplastic; anticonvulsant; antiallergic; immunosuppressive;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antinaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy. The
 CC vectors. The proteins and nucleic acids may be used to treat cancers;
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 280 AA:

Query Match 38.4%; Score 328; DB 21; Length 280;
 Best Local Similarity 67.1%; Pred. No. 8.2e-24;
 Matches 57; Conservative 15; Mismatches 13; Indels 0; Gaps 0;

OY 1 DSAFSPFVDTFAPGSMIIKHPDSTMKERIKNDYOSIEELKDNFKIMCTNAMY 60
 DB 81 ophgftafvdtalapygsmilkpmdfgtmkdkivaneysvtekdafkilmcdnamy 140
 OY 61 NKPEIYYAKAKLHSGMKILSOE 85
 DB 141 nrdptlvvykklakkllhagfkmmksq 165

RESULT 8

AAM38835
 ID AAM38835 standard; Protein; 597 AA.
 XX
 AC AAM38835;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 1980.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 XX
 PR 25-APR-2000; 2000US-0552317.
 XX
 PR 09-JUL-2000; 2000US-0598042.
 XX
 PR 19-JUL-2000; 2000US-0620112.
 XX
 PR 03-NOV-2000; 2000US-0653450.
 XX
 PR 14-SEP-2000; 2000US-0662191.
 XX
 PR 19-OCT-2000; 2000US-0693036.
 XX
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 XX
 DR N-PSDB: AAI57991.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 XX
 PT such as central nervous system injuries -
 XX
 PS Example 3; SEQ ID NO 1980; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AA42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC activating/inhibiting activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 597 AA:

Query Match 38.4%; Score 328; DB 22; Length 597;
 Best Local Similarity 67.1%; Pred. No. 2.1e-23;
 Matches 57; Conservative 15; Mismatches 13; Indels 0; Gaps 0;

OY 1 DSAFSPFVDTFAPGSMIIKHPDSTMKERIKNDYOSIEELKDNFKIMCTNAMY 60
 DB 81 ophgftafvdtalapygsmilkpmdfgtmkdkivaneysvtekdafkilmcdnamy 140

0y	61	NKPEETIYYKAKRKLHSGMKIISQE	85
Db	216	nrptvyykiakklhagfkmskq	240
RESULT 9			
AAAM0621	9		
ID	AAAM0621	standard; Protein; 605 AA.	
AC	AAAM0621:		
DE	22-OCT-2001	(first entry)	
XX			
XX	Human polypeptide SEQ ID NO 5552.		
KW	Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokine; thrombolytic; drug screening; arthritis; inflammation; leukaemia.		
XX			
OS	Homo sapiens.		
PN	WO200153312-A1.		
XX			
PD	26-JUL-2001.		
XX			
PF	26-DEC-2000; 2000MO-US34263.		
XX			
PR	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
PR	29-NOV-2000; 2000US-0727344.		
XX			
XX	(HXSE-) HXSEQ INC.		
PA			
XX			
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
PI	Wang Z, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QA, Zhou P, Goodrich R, Dirmacac RT;		
XX			
DR	WPI: 2001-AA2253/47.		
XX			
XX	N-PSDB: AA159777.		
PT			
PT	Novel nucleic acids and polypeptides, useful for treating disorders		
XX	such as central nervous system injuries -		
XX			
XX	Example 2; SEQ ID NO 5552; 10078bp; English.		
CC			
CC	The invention relates to human nucleic acids (AA157798-AA161369) and		
CC	the encoded polypeptides (AAAM38642-AAAM42213) with neotropic,		
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful		
CC	in gene therapy. A composition containing a polypeptide or polynucleotide		
CC	of the invention may be used to treat diseases of the peripheral nervous		
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		
CC	localised neuropathies and central nervous system diseases, such as		
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		
CC	lateral sclerosis and Shy-Drager Syndrome. Other uses include the		
CC	utilisation of the activities such as: Immune system suppression,		
CC	Activin/inhibin activity, chemotactic/chemokine activity, haemostatic		
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,		
CC	assays for receptor activity, arthritis and inflammation, leukaemias and		
CC	C.N.S disorders		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification.		
XX			
XX	Sequence 605 AA;		

```

Query Match          38.4%; Score 328; DB 22; Length 605;
Best Local Similarity 67.1%; Pred. No. 2.2e-23;
Matches 57; Conservative 15; Mismatches 13; Indels 0; Gaps 0;

OY      1 DPSAFSPVPDVFIAFGYSMIIRHMDFSYMKERIKRINDYOISIELKDNFKLMCTNAMTY 60
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      164 dphgfafprvcdalagysmlikhpmdgtunkdkrtvaneyksvfekfdkflmchdnantyy 223
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

OY      61 NKPEITYYKAARKLILHSGMKTLSQE 85
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      224 nripdlvykkllakklilhagfkmmxkg 248

RESULT 10
ABB63028
ID      ABB63028 standard; Protein: 861 AA.
AC      ABB63028;
XX
XX      26-MAR-2002 (first entry)
DT
DE      Drosophila melanogaster polypeptide SEQ ID NO 15876.
XX
KW      Drosophila; developmental biology; cell signalling; insecticide;
KW      pharmaceutical.
XX
OS      Drosophila melanogaster.
PN      WO200171042-A2.
XX
PD      27-SEP-2001.
XX
PF      23-MAR-2001; 2001WO-US09231.
XX
PR      23-MAR-2000; 2000US-191637P.
XX      11-JUL-2000; 2000US-0614150.
XX
PA      (PEKE ) PE CORP NY.
XX
PI      Venter JC, Adams M, Li PWD, Myers EW;
DR      WPI: 2001-6556860/75.
DR      N-PSTDB; ABL07131.
XX
PT      New isolated nucleic acid detection reagent for detecting 1000 or more
PT      genes from Drosophila and for elucidating cell signalling and cell-cell
PT      interactions -
XX
PS      Disclosure; SEQ ID NO 15876; 21pp + Sequence Listing; English.
XX
CC      The invention relates to an isolated nucleic acid detection reagent
CC      capable of detecting 1000 or more genes from Drosophila. The invention is
CC      useful in developmental biology and in elucidating cell signalling and
CC      cell-cell interactions in higher eukaryotes for the development of
CC      insecticides, therapeutics and pharmaceutical drugs. The invention
CC      sequences genomic DNA sequences (AB101840-AB1616175) and the encoded proteins
CC      (ABBS7737-ABBS72072).
CC      The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ      Sequence 861 AA;

Query Match          33.3%; Score 284.5; DB 22; Length 861;
Best Local Similarity 41.0%; Pred. No. 5.8e-19;
Matches 68; Conservative 30; Mismatches 33; Indels 15; Gaps 4;

OY      1 DSAFFSFPVTDVIAFGYSMIIRHMDFSYMKERIKRINDYOISIELKDNFKLMCTNAMTY 60
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

Db 295 dphgfawpvtddmapyyslstrpmdfstrmqkiddheyaltefddfkImcenalky 354
QY 61 NKPEITYYKAARKLLHSGMKILISOERI-QSLKOSIDFMADQKTRKOKDGTDSQSGHDG 119
Db 355 nhvdtvynkaakrlllygmklhpenlmrslklsygmrel---tarelgfelsnd--- 408
QY 120 GCMQREHEDSGD-----AEAHAFKSPSKENKKDKMDLEDKFSKN 159
Db 409 --msrenndasadegastgtaeeptrpaqlleeerktlrlenapth 452

RESULT 11

AAB63767
ID AAB63767 standard; Protein; 414 AA.

AC AAB63767;

DT 26-MAR-2001 (first entry)

DE Human prostate cancer associated antigen protein sequence SEQ ID NO:1129.

KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;

KW cancer associated antigen; cytostatic; cancer vaccine.

OS Homo sapiens.

PN WO200073801-A2.

PD 07-DEC-2000.

PF 26-MAY-2000; 2000WO-US14749.

PR 28-MAY-1999; 99US-0136526.

PR 10-SEP-1999; 99US-0153454.

PA (LUDWIG) LUDWIG INST CANCER RES.

PI Obata Y;

DR WPI; 2001-025274/03.

PT Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.,
PT cancer -

PS Example 1; Page 708-709; 799pp; English.

CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterized by expression of an abnormal amount of a protein,
CC e.g. cancer.

XX Sequence 414 AA;

Query Match 20.4%; Score 174; DB 22; Length 414;
Best Local Similarity 38.4%; Pred. No. 1.3e-08;

Matches 33; Conservative 22; Mismatches 31; Indels 0; Gaps 0;

QY 1 DSAFSPFVPTDFIAGYSMTIKHPMDFSTMKERIKNDYOSIEELKDNFKIMCTNAMIY 60
Db 134 dparifagpvsikvpydyldhikhpmdfatmrkrlaaggyknlhfeefddllidnckmy 193

QY 61 NKPEITYYKAARKLLHSGMKILISOER 86

Db 194 nardlvfyraavrlrtdggvavltqr 219

RESULT 12

ABG12634
ID ABG12634 standard; Protein; 715 AA.

AC ABG12634;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #12625.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS76821.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID NO 42993; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 715 AA;

Query Match 20.4%; Score 174; DB 22; Length 715;
Best Local Similarity 38.4%; Pred. No. 2.5e-08;

Matches 33; Conservative 22; Mismatches 31; Indels 0; Gaps 0;

QY 1 DSAFSPFVPTDFIAGYSMTIKHPMDFSTMKERIKNDYOSIEELKDNFKIMCTNAMIY 60
Db 108 dparifagpvsikvpydyldhikhpmdfatmrkrlaaggyknlhfeefddllidnckmy 167

QY 61 NKPEITYKAAKKLLHSGMKILSOER 86
 Db 168 nardtlyfyaavrlrdggvylrqar 193

RESULT 13

AAM39231
 ID AAM39231 standard; Protein: 1058 AA.

XX AAM39231;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 2376.

XX Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

XX Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

DR N-PSDB; AAI58387.

XX WPI; 2001-442253/47.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX N-PSDB; AAI58387.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX N-PSDB; AAI58387.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX N-PSDB; AAI58387.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;

Query Match 20.4%; Score 174; DB 22; Length 1058;
 Best Local Similarity 38.4%; Pred. No. 4.2e-08;
 Matches 33; Conservative 22; Mismatches 31; Indels 0; Gaps 0;

QY 1 DPSAFSEFPTWDFAPGYSMIKHPMDFSTMKERIKNNDYQSIETLKNFKMCTNAMY 60
 Db 582 dparifqpyvslkevpdyldhikpmdfatmrkleagynknhefeedldliddcmky 641

QY 61 NKPEITYKAAKKLLHSGMKILSOER 86

Db 642 nardtlyfyaavrlrdggvylrqar 667

RESULT 14

AAM41017
 ID AAM41017 standard; Protein: 1109 AA.

XX AAM41017;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 5948.

XX Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

XX Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

DR N-PSDB; AAI60173.

XX WPI; 2001-442253/47.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX N-PSDB; AAI60173.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX N-PSDB; AAI60173.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 15:50:09; Search time 22.7 Seconds

(without alignments)
175.391 Million cell updates/sec

Title: US-09-687-230-2_COPY_151_313

Perfect score: 855
Sequence: 1 DSAFFSPVDTFIAPGYSM.....NKKKDKMLEDKFSNNLER 163

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	855	100.0	589	4	US-08-942-008-2
2	145.5	17.0	1872	1	US-08-188-582-14
3	145.5	17.0	1872	1	US-08-646-715-14
4	145.5	17.0	1893	1	US-08-188-582-11
5	145.5	17.0	1893	1	US-08-646-715-11
6	129	15.1	2441	3	US-08-194-468-2
7	129	15.1	2441	3	US-08-961-739-2
8	123	14.4	65	1	US-08-227-536-5
9	123	14.4	65	5	PCR-US95-04682-5
10	122	14.3	2414	1	US-08-227-536-2
11	122	14.3	2414	5	PCR-US95-04682-2
12	118	13.8	65	1	US-08-227-536-6
13	118	13.8	65	5	PCR-US95-04682-6
14	108	12.6	65	1	US-08-227-536-7
15	108	12.6	65	5	PCR-US95-04682-7
16	103	12.0	65	1	US-08-227-536-8
17	103	12.0	65	5	PCR-US95-04682-8
18	98	11.5	65	1	US-08-227-536-4
19	98	11.5	65	5	PCR-US95-04682-4
20	94	11.0	65	1	US-08-227-536-3
21	94	11.0	65	5	PCR-US95-04682-3
22	91.5	10.7	976	4	US-08-894-997-50
23	80	9.4	562	2	US-08-973-675-2
24	79.5	9.3	914	1	US-08-487-890A-11
25	79.5	9.3	914	2	US-08-478-435-11
26	79.5	9.3	914	2	US-08-337-483-11
27	79.5	9.3	914	2	US-08-478-373-11

28	79.5	9.3	914	3	US-08-474-671-11	Sequence 11, Appl
29	79.5	9.3	914	3	US-08-483-577A-11	Sequence 11, Appl
30	79.5	9.3	914	4	US-08-897-428-11	Sequence 11, Appl
31	79.5	9.3	914	4	US-08-637-654-11	Sequence 11, Appl
32	79	9.2	676	2	US-08-398-590A-40	Sequence 40, Appl
33	79	9.2	676	4	US-08-894-997-40	Sequence 40, Appl
34	77.5	9.1	2482	1	US-08-328-254-6	Sequence 6, Appl
35	77.5	9.1	3248	1	US-08-353-700-1	Sequence 1, Appl
36	77.5	9.1	3248	5	PCT-US95-16216-1	Sequence 20, Appl
37	76.5	8.9	356	4	US-09-308-003-20	Sequence 1, Appl
38	76	8.9	1618	1	US-07-853-913-4	Sequence 4, Appl
39	75.5	8.8	258	2	US-08-602-941-3	Sequence 3, Appl
40	75.5	8.8	258	3	US-08-961-264-3	Sequence 3, Appl
41	75	8.8	608	2	US-08-736-770-1	Sequence 1, Appl
42	74.5	8.7	576	2	US-08-533-306A-2	Sequence 2, Appl
43	74.5	8.7	576	2	US-08-742-923A-2	Sequence 2, Appl
44	74.5	8.7	816	2	US-08-533-306A-6	Sequence 6, Appl
45	74.5	8.7	816	2	US-08-742-923A-6	Sequence 6, Appl

ALIGNMENTS

```
RESULT 1
US-08-942-008-2
; Sequence 2, Application US/08942008
; Patent No. 6133419
;
; GENERAL INFORMATION:
; APPLICANT: Braselmann, Sylvia
; TITLE OF INVENTION: Nucleotide Sequences that Encode
; TITLE OF INVENTION: Phosphatidylinositol 3' Kinase Associated Proteins and
; TITLE OF INVENTION: Uses Thereof
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: ONYX Pharmaceuticals, Inc.
; STREET: 3031 Research Drive
; CITY: Richmond
; STATE: CA
; COUNTRY: USA
; ZIP: 94806
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,008
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Giotta, Gregory
; REGISTRATION NUMBER: 32,028
; REFERENCE/DOCKET NUMBER: ONYX1027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 262-8710
; TELEFAX: (510) 222-9758
; INFORMATION FOR SEQ. ID NO.: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 589 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-942-008-2

Query Match 100.0%; Score 855; DB 4; Length 589;
Best Local Similarity 100.0%; Pred. No. 2.6e-85;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSAFFSPVDTFIAPGYSMIIKHPMDFSTMKIKIKNDYOSIEELKDNFKIKCTNAMTY 60
DB 151 DSAFFSPVDTFIAPGYSMIIKHPMDFSTMKIKIKNDYOSIEELKDNFKIKCTNAMTY 210
```

QY 61 NRPEITYYKAKKLLHSGMKILISOERISQSLKOSIDFMADLOKTRKOKDGTDSQSGEDG 120
DB 211 NRPEITYYKAKKLLHSGMKILISOERISQSLKOSIDFMADLOKTRKOKDGTDSQSGEDG 270
QY 121 CWREREDSGDAFAHAFKSPSKENKKDKMDLDRKSNLRLR 163
DB 271 CWREREDSGDAFAHAFKSPSKENKKDKMDLDRKSNLRLR 313

RESULT 2

US-08-188-582-14
; Sequence 14, Application US/08188582
; Patent No. 5534410

GENERAL INFORMATION:

APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dylact, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TATS AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:
LENGTH: 1872 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-188-582-14

Query Match 17.0%; Score 145.5; DB 1; Length 1872;
Best Local Similarity 26.0%; Pred. No. 6.3e-07;
Matches 45; Conservative 33; Mismatches 72; Indels 23; Gaps 4;

QY 6 FSPVYDFLAPGYSMIKHPMDSYTKKIKNDYOSIELKDNFKLMCTNMYKPPET 65
DB 1528 FHPVKKKFPDYKYIVNPMDELIRKNISKKHKYQSRSEFLDDVNLILANSVKYGPES 1587
QY 66 IYKAKKLLHSGMKILSQ--ERISQSLK-----QSIDFMADLOKTRKOK 107
DB 1588 QYTKIAQELVNCYQTLFEYDEHLTQLEKDICTAKAALAEALFESLDPTGPPYPPQP 1647
QY 108 DGTDTQS---GDDGCGWREREDS--GDAEAHAFKSPSKENKKDKMDLEDR 155

DB 1648 DLYDTNLSMSRDSAYFODESNMAYLDIPSATPEKQVYQEGBDGGLADEE 1700

RESULT 3

US-08-646-715-14
; Sequence 14, Application US/08646715
; Patent No. 5637686

GENERAL INFORMATION:

APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dylact, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TATS AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,715
FILING DATE: 09-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/188,582
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:
LENGTH: 1872 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-646-715-14

Query Match 17.0%; Score 145.5; DB 1; Length 1872;
Best Local Similarity 26.0%; Pred. No. 6.3e-07;
Matches 45; Conservative 33; Mismatches 72; Indels 23; Gaps 4;

QY 6 FSPVYDFLAPGYSMIKHPMDSYTKKIKNDYOSIELKDNFKLMCTNMYKPPET 65
DB 1528 FHPVKKKFPDYKYIVNPMDELIRKNISKKHKYQSRSEFLDDVNLILANSVKYGPES 1587
QY 66 IYKAKKLLHSGMKILSQ--ERISQSLK-----QSIDFMADLOKTRKOK 107
DB 1588 QYTKIAQELVNCYQTLFEYDEHLTQLEKDICTAKAALAEALFESLDPTGPPYPPQP 1647
QY 108 DGTDTQS---GDDGCGWREREDS--GDAEAHAFKSPSKENKKDKMDLEDR 155
DB 1648 DLYDTNLSMSRDSAYFODESNMAYLDIPSATPEKQVYQEGBDGGLADEE 1700

Matches 23; Conservative 6; Mismatches 19; Indels 0; Gaps 0;

QY 16 PGSMIIKHPMDFSTMKERIKNDYOSIELKDNFKLMCTNMIYKNP 63
 18 PDYHKIITKQPMDMGTIKRLENNYTWASSECMDFTNTFTNCIYIYKNP 65

RESULT 9
 PCT-US95-04682-5
 Sequence 5, Application PC/TUS9504682

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION
 TITLE OF INVENTION: FACTOR P300 AND USES OF P300

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
 STREET: Ten Post Office Square
 CITY: Boston
 STATE: MA

COUNTRY: US

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/04682

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/227,536

FILING DATE: 14-April-1994

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Holliday C. Heine, Ph.D.

REGISTRATION NUMBER: 34,346

REFERENCE/DOCKET NUMBER: DECI-308X999

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-2290

TELEFAX: (617) 451-0313

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 65 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: Internal

PCT-US95-04682-5

Query Match 14.4%; Score 123; DB 5; Length 65;
 Best Local Similarity 47.9%; Pred. No. 1.6e-06;
 Matches 23; Conservative 6; Mismatches 19; Indels 0; Gaps 0;

QY 16 PGYSMIKHPMDFSTMKERIKNDYOSIELKDNFKLMCTNMIYKNP 63
 18 PDYHKIITKQPMDMGTIKRLENNYTWASSECMDFTNTFTNCIYIYKNP 65

RESULT 10
 US-08-227-536-2

Sequence 2, Application US/08227536

Patent No. 5658784

GENERAL INFORMATION:

APPLICANT: Eckner, Richard

APPLICANT: Even, Mark

APPLICANT: Livingston, David

TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION

TITLE OF INVENTION: FACTOR P300 AND USES OF P300

NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
 STREET: Ten Post Office Square
 CITY: Boston
 STATE: MA

COUNTRY: US

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/227,536

FILING DATE: 14-Apr-1994

CLASSIFICATION: 436

ATTORNEY/AGENT INFORMATION:

NAME: Williams Ph.D., Kathleen A.

REGISTRATION NUMBER: 34,380

REFERENCE/DOCKET NUMBER: DECI-308XX

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-2290

TELEFAX: (617) 451-0313

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2414 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-227-536-2

Query Match 14.3%; Score 122; DB 1; Length 2414;
 Best Local Similarity 34.3%; Pred. No. 0.00033;
 Matches 34; Conservative 15; Mismatches 42; Indels 8; Gaps 4;

QY 1 DPGAF-ESFPYTDFT--APGYSMIKHPMDFSTMKERIKNDYOSIELKDNFKLMCTNMIYKNP 57
 1069 DPESLPFRQPVDPOLGIDPDIKSPMDLSTIRKIDTGOYQBPWOYVDIVIMFNNA 1128

QY 58 MYNKPETIYYKAAKILHSGMKILSOERISLKSIDF 96

DB 1129 WLYNKRKTSKVYKSKL-----SEVEEQE-IDPVMSLQY 1162

RESULT 11
 PCT-US95-04682-2
 Sequence 2, Application PC/TUS9504682

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION
 TITLE OF INVENTION: FACTOR P300 AND USES OF P300

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
 STREET: Ten Post Office Square
 CITY: Boston
 STATE: MA

COUNTRY: US

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/04682

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/227,536

FILING DATE: 14-April-1994

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Holliday C. Heine, Ph.D.
REGISTRATION NUMBER: 34,346
REFERENCE/DOCKET NUMBER: DFCI-308Xq999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-2290
TELEFAX: (617) 451-0313
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2414 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-04682-2

Query Match 14.3%; Score 122; DB 5; Length 2414;
Best Local Similarity 34.3%; Pred. No. 0.00033;
Matches 34; Conservative 15; Mismatches 42; Indels 8; Gaps 4;

QY 1 DPSAF-FSEPTDFT--APGYSMIHKHMDSTMEKIKNDYQSTIEELKDNFKMCTNMA 57
DB 1069 DPESEFPQPDPLGLGIPDFIDYKSPMDLSTIKRLDGTQYQEPQYVDIWMFNNA 1128
QY 58 MIYNKPEIYYKAAKILHSGKILSQERISQISKQIDF 96
DB 1129 WLYNRKTSRYVKKYCKSL---SEVEPEGR-IDPVNQSLCY 1162

RESULT 12
US-08-227-536-6
Sequence 6, Application US/08227536
Patent No. 5658784
GENERAL INFORMATION:
APPLICANT: Eckner, Richard
APPLICANT: Livingston, David
TITLE OF INVENTION: NUCLEIC ACID, ENCODING TRANSCRIPTION
TITLE OF INVENTION: FACTOR P300 AND USES OF P300
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
STREET: Ten Post Office Square
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/227,536
FILING DATE: 14-APR-1994
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Williams Ph.D., Kathleen A.
REGISTRATION NUMBER: 34,380
REFERENCE/DOCKET NUMBER: DFCI-308XX
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-2290
TELEFAX: (617) 451-0313
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 65 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal

US-08-227-536-6

Query Match 13.8%; Score 118; DB 1; Length 65;
Best Local Similarity 47.8%; Pred. No. 5.8e-06;
Matches 22; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY 18 YSMIHKHMDSTMEKIKNDYQSTIEELKDNFKMCTNMAIYNKP 63
DB 20 YHDIHKPMDLSTVKKRMENRDYRDAQEPADYRLMFSNCTYNNP 65

RESULT 13
PCT-US95-04682-6
Sequence 6, Application PC/TUS9504682
GENERAL INFORMATION:

APPLICANT:
TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION
TITLE OF INVENTION: FACTOR P300 AND USES OF P300
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
STREET: Ten Post Office Square
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04682
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/227,536
FILING DATE: 14-April-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Holliday C. Heine, Ph.D.
REGISTRATION NUMBER: 34,346
REFERENCE/DOCKET NUMBER: DFCI-308Xq999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-2290
TELEFAX: (617) 451-0313
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 65 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
PCT-US95-04682-6

Query Match 13.8%; Score 118; DB 5; Length 65;
Best Local Similarity 47.8%; Pred. No. 5.8e-06;
Matches 22; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY 18 YSMIHKHMDSTMEKIKNDYQSTIEELKDNFKMCTNMAIYNKP 63
DB 20 YHDIHKPMDLSTVKKRMENRDYRDAQEPADYRLMFSNCTYNNP 65

RESULT 14
US-08-227-536-7
Sequence 7, Application US/08227536
Patent No. 5658784
GENERAL INFORMATION:

APPLICANT: Eckner, Richard
APPLICANT: Even, Mark
TITLE OF INVENTION: NUCLEIC ACID, ENCODING TRANSCRIPTION
TITLE OF INVENTION: FACTOR P300 AND USES OF P300
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
STREET: Ten Post Office Square
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/227,536
FILING DATE: 14-Apr-1994
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Williams Ph.D., Kathleen A.
REGISTRATION NUMBER: 34,380
REFERENCE/DOCKET NUMBER: DFCI-308XX
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-2290
TELEFAX: (617) 451-0313
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 65 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
US-08-227-536-7

Query Match 12.6%; Score 108; DB 1; Length 65;
Best Local Similarity 36.0%; Pred. No. 7.2e-05;
Matches 18; Conservative 15; Mismatches 17; Indels 0; Gaps 0;

QY 12 DFLAPGYSMTIKHPDEFSTMKERIKNDYQSTIELKDNFKLMCTNMIYN 61
DB 14 ELXPEYELIRKPVDFKRIKERIRNHKYSIGDLEKDVMLCHNAQTFN 63

RESULT 15
PCT-US95-04682-7
Sequence 7, Application PC/TUS9504682
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION
TITLE OF INVENTION: FACTOR P300 AND USES OF P300
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
STREET: Ten Post Office Square
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04682
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/227,536
FILING DATE: 14-April-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Holliday C. Heine, Ph.D.
REGISTRATION NUMBER: 34,346
REFERENCE/DOCKET NUMBER: DFCI-308Xg999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-2290
TELEFAX: (617) 451-0313
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 65 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
PCT-US95-04682-7

Query Match 12.6%; Score 108; DB 5; Length 65;
Best Local Similarity 36.0%; Pred. No. 7.2e-05;
Matches 18; Conservative 15; Mismatches 17; Indels 0; Gaps 0;

QY 12 DFLAPGYSMTIKHPDEFSTMKERIKNDYQSTIELKDNFKLMCTNMIYN 61
DB 14 ELXPEYELIRKPVDFKRIKERIRNHKYSIGDLEKDVMLCHNAQTFN 63

Search completed: July 11, 2002, 15:50:11
Job time: 549 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 15:50:43 ; Search time 29.85 Seconds

(without alignments)
524.709 Million cell updates/sec

Title: US-09-687-230-2_COPY_151_313

Perfect score: 855
Sequence: 1 DPSAFSEFPVTDFTIAPGYSM.....NKKKDKMLDEKRSNNLER 163

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	200.5	23.5	636	2 T18845	hypothetical prote
2	174	20.4	715	2 T12534	hypothetical prote
3	169	19.8	1214	2 JC2069	zinc-finger protei
4	163.5	19.1	556	2 D96791	hypothetical prote
5	158.5	18.5	1572	2 S45251	SNF2alpha protein
6	158.5	18.5	1586	2 S39580	HRM protein - hum
7	151.5	17.7	757	2 S68142	probable transcrip
8	148.5	17.4	454	2 T37933	transcription acti
9	146.5	17.1	733	2 T28145	RING3 kinase - chl
10	146	17.1	2068	2 A47371	transcription init
11	145.5	17.0	1893	2 A40262	transcription init
12	144	16.8	1022	2 T53078	homeotic gene regu
13	144	16.8	1613	2 S39059	protein Brg1 - hum
14	144	16.8	1647	2 S45252	SNF2beta protein -
15	142	16.6	754	2 A56619	female sterile hom
16	140	16.4	374	2 T33328	hypothetical prote
17	135.5	15.8	1865	2 A43742	transcription init
18	135	15.8	2038	2 A43742	female sterile hom
19	133.5	15.6	578	2 T40984	transcription fact
20	132	15.4	1450	2 S32373	DNA-binding protei
21	131.5	15.4	832	1 S71788	P/CAP protein - hu
22	130.5	15.3	638	2 S67605	hypothetical prote
23	130	15.2	439	2 S28051	transcription fact
24	129	15.1	2441	2 S39161	CREB-binding prote
25	128	15.0	1633	2 JC5056	polyoma 1 - chic
26	124.5	14.6	1680	2 T41628	probable transcrip
27	124	14.5	1638	2 A42091	transcription acti
28	123	14.4	2440	2 S39162	transcription coac
29	122	14.3	1879	2 T19481	hypothetical prote

30	122	14.3	2414	2 A54277	transcription adap
31	120	14.0	542	2 S54260	bromodomain protei
32	118	13.8	449	2 T12495	hypothetical prote
33	117	13.7	1332	2 S41552	probable transcrip
34	115.5	13.5	979	2 T40006	hypothetical prote
35	115	13.5	1479	2 T17401	transcription regu
36	114	13.3	361	2 T42517	bromodomain protei
37	112	13.1	400	2 T00472	probable RING3 pro
38	111.5	13.0	369	2 T46098	hypothetical prote
39	111	13.0	586	2 T47620	histon acetyltrans
40	110.5	12.9	1474	2 T20488	hypothetical prote
41	110	12.9	1251	2 B86194	hypothetical prote
42	110	12.9	2027	2 S60123	hypothetical prote
43	110	12.9	2056	2 G88564	protein R10E1.1
44	105.5	12.3	1250	2 T22845	hypothetical prote
45	105.5	12.3	1703	2 S15047	SNF2 protein - yea

ALIGNMENTS

RESULT 1

T18845

hypothetical protein C01H6.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000

C:Accession: T18845

R:Berts, M.

submitted to the EMBL Data Library, April 1996

A:Reference number: Z19030

A:Accession: T18845

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-636 <MIL>

A:Cross-references: EMBL:Z12528; PIDN:CAA95779.1; GSPDB:GN00019; CESP:C01H6.7

A:Experimental source: clone C01H6

C:Genetics:

A:Gene: CESP:C01H6.7

A:Map position: 1

A:Insertions: 20/1; 198/2; 265/3; 451/3; 489/3; 525/3

C:Superfamily: bromodomain homology

F:172-227/Domain: bromodomain homology <BRO>

Query Match

Best Local Similarity 23.5%; Score 200.5; DB 2; Length 636;

Matches 54; Conservative 32; Mismatches 56; Indels 29; Gaps 5;

QY	1	DPSAFSEFPVTDFTIAPGYSMIIRKHPMDESTMKEIKKNDYOSIEELKDNFKLMCTNAMY 60	
DB	167	DPEGYFAFPVTPSKAPYRDIITKTPMDLQITRENIENIGKIASLPAMKEDCELIYSNAFY 226	
QY	61	NKETIYKAKKLLHSGMKILSGERIOSLSKSIDF-----NADLOKTR-- 104	
DB	227	NQPTVTVYLAARKLSNLIAVYFGQYLRFLPHSLPMANKIPFELVIGIRLAPVPERKRTMN 286	
QY	105	KQKDGTTSQSGDEGCGWQ-----REREDSGDAEAFAPSPSENKKKK 149	
DB	287	KRAAVRDMGTSD--CLQVADPKVRERLSAKLPEAN-----NPKNKKMG 330	

RESULT 2

T12534

hypothetical protein DKFZp434B094.1 - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 02-Sep-2000

C:Accession: T12534

R:Wambutt, R.; Heuener, D.; Mewes, H.W.; Gaassenhuber, J.; Wilmann, S.

submitted to the Protein Sequence Database, June 1999

A:Reference number: Z17524

A:Accession: T12534

A>Status: preliminary

A:Molecule type: mRNA


```

RESULT 10
A:Accession: A47371
transcription initiation factor IID 230k chain - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
R:Accession: A47371
R:OKubio, T.; Gong, D.W.; Yamashita, S.; Horikoshi, M.; Roeder, R.G.; Nakatani, Y.
Genes Dev. 7, 1033-1046, 1993
A:Title: Drosophila 230-kD TFIIID subunit, a functional homolog of the human cell cycle
A:Reference number: A47371, MUID:93279463
A:Accession: A47371
A:Status: preliminary
A:Molecule type: mRNA; protein
A:Residues: 1-2068 <KOK>
A:Cross-references: GB:S61883; NID:g385550; PIDN:AA226991.1; PID:g385551
A:Note: sequence inconsistent with nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:133002, NCBIPI:133003)
C:Genetics:
A:Gene: FLYBase:Ta1250
A:Cross-references: FLYBase:Fpagn0010355
C:Superfamily: unassigned bromodomain proteins; bromodomain homology
C:Keywords: transcription initiation
F:1498-1553/Domain: bromodomain homology <BR01>
F:1620-1675/Domain: bromodomain homology <BR0>

Query Match 17.1%; Score 146; DB 2; Length 2068;
Best Local Similarity 31.2%; Pred. No. 0.0013;
Matches 30; Conservative 27; Mismatches 37; Indels 2; Gaps 1;

OY 6 FSPVPVDFIAPGYSMTIKHPMDESTMKEKIKNDYQSIPELKDNEFLMCTNMIYKPER 65
DB 1498 FLEPVSARKKPYDYVYVTKPMDIQTMREYINQRRTSRFPELDIKQIVDNSLTVNGPS 1557
OY 66 IYKAKKILHSGMKILS--QERISLKSQSIDFMAD 99
DB 1558 AYTLAAGRMWSSCFELAEERDKIMRLKAINPLLD 1593

RESULT 11
A40262
transcription initiation factor IID 250k chain splice form 1 - human
N:Alternate names: 210k nuclear DNA-binding cell cycle gene 1 protein (CCG1); TATA-binding
N:Contains: transcription initiation factor IID 250k chain splice form 2
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: A40262; S03005; S00830; S32352; S32353
R:Seikiguchi, T.; Nohiro, Y.; Nakamura, Y.; Hisamoto, N.; Nishimoto, T.
Mol. Cell. Biol. 11, 3317-3325, 1991
A:Title: The human CCG1 gene, essential for progression of the G-1 phase, encodes a 210-
A:Reference number: A40262; MUID:91246200
A:Accession: A40262
A:Molecule type: mRNA
A:Residues: 1-177,199-1893 <SEK>
A:Cross-references: GB:D90359; NID:g559319; PIDN:BA414374.1; PID:g219528
R:Note: nucleotide sequence not complete
R:Seikiguchi, T.; Miyata, T.; Nishimoto, T.
submitted to the EMBL Data Library, February 1988
A:Reference number: S03005
A:Accession: S03005
A:Molecule type: mRNA
A:Residues: 'MYR',60-117,199-1604,'DNECSSKANIVICLYOYSSQIEELRF' <SE5>
A:Cross-references: EMBL:X07024; NID:g29732; PIDN:CAA30073.1; PID:g29733
A:Note: this sequence has been revised in reference A40262
R:Seikiguchi, T.; Miyata, T.; Nishimoto, T.
EMBO J. 7, 1683-1687, 1988
A:Title: Molecular cloning of the cDNA of human x chromosomal gene (CCG1) which compleme
A:Reference number: S00830; MUID:89005056
A:Accession: S00830
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA

```

A:Residues: 1372-1379;1386-1604, 'DNE', 'CSSKANDVICLIQYCSSQIEELRF' <SE6>
A:Cross-references: EMBL:X07024
R:Ruppert, S.; Wang, E. H.; Tjian, R.
A:Title: Cloning and expression of human TAF(II)250: a TBP-associated factor implicated in
Nature 362, 175-179, 1993
A:Reference number: S32352
A:Accession: S32352
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 178-198 <RUP>
R:Hisatke, K.; Hasegawa, S.; Takada, R.; Nakatani, Y.; Horikoshi, M.; Roeder, R. G.
Nature 362, 179-181, 1993
A:Title: The p250 subunit of native TATA box-binding factor TFIID is the cell-cycle r
A:Reference number: S32353; MUID:93196705
A:Accession: S32353
A:Status: preliminary
A:Molecule type: protein
A:Residues: 'P';587-595;1009-1022;1351-1355;1357-1360 <HIS>
A:Note: 1351-VAL, 1353-LYS, and 1354-GLU were also found
C:Genetics:
A:Gene: GDB:TAF2A; CCG1; CCGS; NSCL2; TAFII250
A:Cross-references: GDB:120573; OMIM:113650
A:Map position: Xq13.1-Xq13.1
C:Superfamily: transcription initiation factor IID 250K chain; bromodomain homology;
C:Keywords: alternative splicing; cell cycle control; DNA binding; duplication; phosph
F:1-1893/Product: transcription initiation factor IID 250K chain splice form 1 stat
F:1-177/199-1893/Product: transcription initiation factor IID 250K chain splice form
F:1216-1295/Domain: HMG box homology <HMG1>
F:1372-1379/Region: nuclear location signal
F:1426-1481/Domain: bromodomain homology <BR01>
F:1549-1604/Domain: bromodomain homology <BR02>
F:1618,1740,1751,1847,1871/Binding site: phosphate (Ser) (covalent) (by casein kinase
F:1678,1054,1684/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kin
F:1020,1361/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) stat
F:1381,1400/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase)

THIS PAGE BLANK (USPTO)


```
Query Match Similarity      20.6%; Score 176; DB 1; Length 1214;
Best Local Similarity       40.4%; Pred. No. 7,3e-07;
Matches    36; Conservative   20; Mismatches   33; Indels     0; Gaps      0;
```

OY 1 DPSAFSPDYDTFIAPGYSMIIKHPMDSTYMEKIKNNDOYSIEELKNFKIMCTNAMLY 60
|| : || | | | | | | | | | | | | | | | | | | | |
Db DPAHFAEAVNLEVPDYLEFISKPMDESTMRRKLSHLFTLEEFEEDFNLIYNCKMY 677

OY 61 NKEPIYYKAANKKLHSKGKILISOERIGS 89
| : ||:::|| : | | | | | : | :
Db 678 NAKDTIFHRAAYRLNDLGGAILRHARROA 706

RESULT 2
BRDL_HUMAN STANDARD; PRT; 1058 AA.
ID BRDL_HUMAN O95696;
AC O95696;
CD 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bromodomain-containing protein 1 (BRL40-like protein).
DR BRL1 OR BRL OR BRPF2..
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20071128; PubMed=10602503;
RA McCullagh P., Chaplin T., Meerabux J., Grenzelias D., Lillington D.,
RA Poulsom R., Gregorin A., Saha V., Young B.D.;
RT "The cloning, mapping and expression of a novel gene, BRL, related to
RT the Afrl leukemia gene.";
RL Oncogene 18:7442-7452(1999).
[2]
SEQUENCE FROM N.A.
RA Hunt A.;
RL Submitted (Aug-1998) to the EMBL/Genbank/DDBJ databases.

CC -1 SUBCELLULAR LOCATION: Nuclear.
CC -1 TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS.
CC -1 SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -1 SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC -1 SIMILARITY: CONTAINS 1 PMWP DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb.sib.ch).

Cc EMBL; AF005067; AACF34320.1; -
Dr EMBL; Z98885; CAB1574.1; -.
Dr MIM: 604589; -
Dr InterPro: IPR001487; Bromodomain.
Dr InterPro: IPR001965; PHD.
Dr InterPro: IPR000313; PMWP.
Dr Pfam; PF00439; bromodomai; 1.
Dr Pfam; PF00628; PHD; 1.
Dr Pfam; PF00855; PMWP; 1.
Dr PRINTS; PR00503; BROMODOMAIN.
Dr SMART; SM00297; BROMO; 1.
Dr SMART; SM00249; PHD; 2.
Dr SMART; SM00293; PMWP; 1.
Dr PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
Dr PROSITE; PSS0014; BROMODOMAIN_2; 1.
Dr PROSITE; PSS0812; PMWP; 1.
KW Nuclear protein; zinc-finger; Bromodomain.
KM ZN_FING 214
TT ZN_FING 214 264
TD PHD-TYPE.
TM DOMAIN 579 649 BROMODOMAIN.

```

FR DOMAIN      929   1012    PMWP.
SQ SEQUENCE    1058 AA; 119519 MW; 6E7B07E8A030E104 CRC64;

Query Match          20.4%; Score 174; DB 1; Length 1058;
Best Local Similarity 38.4%; Pred. No. 9.1e-07;
Matches 33; Conservative 22; Mismatches 31; Indels 0; Gaps 0;

QY      1 DPSAFSPVPTDFIAPGSMITKHMDRSTMEKTKNNDYOSIEELKNFKMTNNTMAY 60
       |||::|||::| | |||||::|::|::|::|::|::|::|::|::|::|::|::|
DB      562 DPARIFAQPVSLKEPVDYIDHIKHMDRIATMKRLLEAGYIKNLHFEEDFDLIDNCMY 641
QY      61 NKPEITYYKAANKILHSCKILSOER 86
DB      642 NARDTVFYRAAVRLRDGGGVYLRRAR 667

RESULT      3
BRFL_HUMAN BRFL_HUMAN STANDARD; PRT; 1214 AA.
AC         P55201; Q9UHI0;
DT         01-OCT-1996 (Rel. 34, Created)
DT         01-OCT-1996 (Rel. 34, Last sequence update)
DT         16-OCT-2001 (Rel. 40, Last annotation update)
DE         Peregrin (Bromodomain and PHD finger-containing protein 1) (BR140
           protein).
GN         BRPF1 OR BR140.
OS         Homo sapiens (Human).
OC         Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX         Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX         NCBI_TaxID=9606;
RN         [1]
RP         SEQUENCE FROM N.A.
RX         MEDLINE=94161726; PubMed=7906940;
RA         Thompson K.A., Wang B., Argraues W.S., Giaccotti F.G., Schranck D.P.,
RA         Rusalski E.;
RT         "BR140, a novel zinc-finger protein with homology to the TAF250
           subunit of TFIIID."
RL         Biochem. Biophys. Res. Commun. 198.1143-1152(1994).
RN         [2]
RP         SEQUENCE FROM N.A.
RX         Hu S.N., Dong W., Zeng Y.X., Yu J., Yang H.M.;
RL         Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
CC         -1- FUNCTION: UNKNOWN. POSSIBLE TRANSCRIPTION ACTIVATOR.
CC         -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC         -1- TISSUE SPECIFICITY: HIGH LEVELS IN TESTIS.
CC         -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC         -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC         -1- SIMILARITY: CONTAINS 1 PMWP DOMAIN.
-----
CC         This SWISS-PROT entry is copyright. It is produced through a collaboration
CC         between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC         the European Bioinformatics Institute. There are no restrictions on its
CC         use by non-profit institutions as long as its content is in no way
CC         modified and this statement is not removed. Usage by and for commercial
CC         entities requires a license agreement (See http://www.isb.ch/announce/
CC         or send an email to license@isb.slb.ch).
-----
DR      EMBL; M91585; AAB02119.1; -.
DR      EMBL; AF176815; AAF19605.1; -.
DR      MIM; 602410; -.
DR      InterPro; IPR001487; Bromodomain.
DR      InterPro; IPR001965; PHD.
DR      InterPro; IPR000313; PMWP.
DR      InterPro; IPR000822; ZnF-C2H2.
DR      Pfam; PF00439; bromodomain; 1.
DR      Pfam; PF00628; PHD; 1.
DR      Pfam; PF00855; PMWP; 1.
DR      PRINTS; PR00503; BROMODOMAIN.
DR      SMART; SM00297; BROMO; 1.
DR      SMART; SM00249; PHD; 2.
DR      SMART; SM00293; PMWP; 1.
DR      SMART; SM00355; ZnF_C2H2; 1.

```


DR PROSITE: PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
DR PROSITE: PS50812; PWM; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein;
KM Zinc-finger; Bromodomain.
FT ZN_FING 21 47 C2H2-TYPE.
FT ZN_FING 273 323 PHD-TYPE.
FT ZN_FING 386 400 C4-TYPE.
FT DOMAIN 645 715 BROMODOMAIN.
FT DOMAIN 1085 1168 PWM.
FT CONFLICT 299 299 E -> A (IN REF. 2).
FT CONFLICT 729 729 V -> L (IN REF. 2).
SQ SEQUENCE 1214 AA; 137542 MW; C530CD2FE3083A53D CRC64;

Query Match 19.8%; Score 169; DB 1; Length 1214;
Best Local Similarity 37.5%; Pred. No. 2,7e-06;
Matches 36; Conservative 19; Mismatches 41; Indels 0; Gaps 0;

QY 1 DPSAFSPVPTDFIAPGYSMIKHPMDFSTMKERIKNDYOSIEELKDNFKIMCTNAMY 60
DB 648 DTGNFSEPVLPSEVPLDHIKKPMDFPMKQMLEAVRYINFDDFEEDFNLYVSNCLKY 707
OY 61 NKPEITYKAKKILHSGMKILSQERIOSLKOSTDF 96
DB 708 NAKDTIFRAAVRLKEGGAVRQARROAKRMGIDF 743

RESULT 4
ID SN22_HUMAN STANDARD; PRT; 1586 AA.
AC P51531;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Possible global transcription activator SNF2L2 (SNF2-alpha).
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94038910; PubMed=8223438;
RA Muchardt C., Yaniv M.;
RT "A human homologue of Saccharomyces cerevisiae SNF2/SMI2 and
RT Drosophila brm genes potentiates transcriptional activation by the
RT glucocorticoid receptor";
RL EMO J. 12:4279-4290(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=94268902; PubMed=8208605;
RA Chiba H., Muramatsu M., Nomoto A., Kato H.;
RT "Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and
RT Drosophila brhma are transcriptional coactivators cooperating with
RT the estrogen receptor and the retinoic acid receptor";
RL Nucleic Acids Res. 22:1815-1820(1994).
CC -1- FUNCTION: TRANSCRIPTIONAL COACTIVATOR COOPERATING WITH NUCLEAR
CC HORMONE RECEPTORS TO POTENTIATE TRANSCRIPTIONAL ACTIVATION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL: X72889; CAAS1407.1; -
CC EMBL: D62155; BAA05142.1; -
CC MIM: 600014; -
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; SNF2_N.
DR Pfam: PF00439; Bromodomain; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00176; SNF2_N; 1.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR PROSITE: PS00633; BROMODOMAIN_1; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
KW Transcription regulation; Nuclear protein; Activator; Bromodomain;
KM ATP-binding; Helicase; Alternative splicing.
FT DOMAIN 216 238 POLY-GLN.
FT DOMAIN 241 249 POLY-GLN.
FT DOMAIN 555 558 POLY-ARG.
FT DOMAIN 639 646 POLY-GLU.
FT NP_BIND 745 752 ATP (POTENTIAL).
FT SITE 847 850 DEGR BOX.
FT DOMAIN 1293 1297 POLY-GLU.
FT DOMAIN 1415 1485 BROMODOMAIN.
FT DOMAIN 1514 1525 POLY-GLU.
FT VARSPLIC 1397 1414 MISSING (IN SHORT ISOFORM).
FT CONFLICT 239 239 P -> POOP (IN REF. 2).
FT CONFLICT 390 390 Q -> E (IN REF. 2).
FT CONFLICT 509 509 G -> S (IN REF. 2).
FT CONFLICT 707 707 W -> R (IN REF. 2).
FT CONFLICT 1135 1135 C -> H (IN REF. 2).
FT CONFLICT 1390 1390 D -> V (IN REF. 2).
SQ SEQUENCE 1586 AA; 180762 MW; FA537E2A2392807A CRC64;

Query Match 18.5%; Score 158.5; DB 1; Length 1586;
Best Local Similarity 30.3%; Pred. No. 2,7e-05;
Matches 43; Conservative 27; Mismatches 47; Indels 25; Gaps 4;

QY 16 PGYSMIKHPMDFSTMKERIKNDYOSIEELKDNFKIMCTNAMYKPEITYYAAKLL 75
DB 1433 PEYELINKPVPDFPKIRKINRHRYSGLDEKVMMLCHNAQGFNLEGSQYEDSYLV- 1491
OY 76 HSGMKILSQERIOSLKOSIDFMADLOKTRKOKDGTDSQSGEDGCMQREDESGDAEAH 135
DB 1492 -----OSV-FKSAROKTAKEESEDSNEE-----EEDEESESSEAK 1529
QY 136 AFKSPSKENKKRDMLEDKR 157
DB 1530 SVKVKIKLKKRDKG--RDKGK 1549

RESULT 5
ID T2D1_DROME STANDARD; PRT; 2068 AA.
AC P51123;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription initiation factor TFIID 230 kDa subunit (TAFII-230)
DE (TAFII250) (TBP-associated factor 230 kDa) (P230).
GN TAF250.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 63-75 AND 540-546.

```

RX MEDLINE-93279463; PubMed-8504928;
RA Kokubo T., Gong D.-W., Yamashita S., Horikoshi M., Roeder R.G.,
RA Nakatani Y.;
RT "Drosophila 230-kD TFIIID subunit, a functional homolog of the human
RT cell cycle gene product, negatively regulates DNA binding of the TAFII
RT box-binding subunit of TFIIID."
RL Genes Dev. 7:1033-1046(1993).
CC -1- FUNCTION: MAY PLAY AN ESSENTIAL ROLE IN TFIIID ASSEMBLY BY
CC INTERACTING WITH BOTH TBP AND OTHER TAF, AS WELL AS SERVING TO
CC LINK THE CONTROL OF TRANSCRIPTION TO THE CELL CYCLE. ESSENTIAL FOR
CC PROGRESSION OF THE G1 PHASE OF THE CELL CYCLE. POSSESSES DNA-
CC BINDING ACTIVITY. IS A NEGATIVE REGULATOR OF THE TATA BOX-BINDING
CC ACTIVITY OF TBP.
CC -1- SUBUNIT: TF2D IS COMPOSED OF TBP AND A VARIETY OF TBP-ASSOCIATED
CC FACTORS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC -1- SIMILARITY: TO HUMAN TAFII-250 (CCG1). SOME TO S.POMBE TAFII-111
CC AND TO S.CEREVISIAE TAF145.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S61883; AAB26991.2; -
DR TRANSFAC; T02119; -
DR FLYBASE; FBgn0010355; Taf250.
DR InterPro; IPR000637; AT_hock.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR001878; ZnF_CCHC.
DR Pfam; PF02178; AT_hock; 1.
DR Pfam; PF00439; Bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00384; AT_hock; 1.
DR SMART; SM00297; BROMO; 2.
DR SMART; SM00343; ZnF_C2HC; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 2.
DR PROSITE; PS50014; BROMODOMAIN_2; 2.
KW Bromodomain; Nuclear protein; DNA-binding; Cell cycle; Repeat;
KW Transcription regulation; Phosphorylation.
FT DNA_BIND 1247 1360 HMG BOX (POTENTIAL).
FT DOMAIN 1445 1451 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 1490 1560 BROMODOMAIN 1.
FT DOMAIN 1612 1682 BROMODOMAIN 2.
FT DOMAIN 1995 2068 GLN-RICH.
FT VARIANT 575 575 P -> S.
SQ SEQUENCE 2068 AA; 232494 MW; AD6A5ABF28B59531 CRC64;

```

```

Query Match 17 1%; Score 146; DB 1; Length 2068;
Best Local Similarity 31.2%; Pred. No. 0.00039;
Matches 30; Conservative 27; Mismatches 37; Indels 2; Gaps 1;

```

```

OY 6 FESPVTDFAPGYSMIIKRPDPSTKRIKNDYQSIIEELKDNFKIMCTNANIKKPEP 65
I |||: | | | | | | | | | | | | | | | | | | | | | | | | | | |
D 1498 FLEPVSARKVPDYRYVTPMDLQTRREYIRQRRTSRREMFLEDLKQIDNSLTVGPOS 1557
I | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1558 AYTIAQRMFSSCFELLAREDKIMLEKAINPLLD 1593
I | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 66 IYKAKKLLHSGMKTL--QERIOSLKOSIDPMAD 99
I | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1558 AYTIAQRMFSSCFELLAREDKIMLEKAINPLLD 1593
I | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

DT 15-JUL-1998 (Rel. 36, last annotation update)
DE Transcription initiation factor TFIIID 250 kDa subunit (TAFII-250)
DE (TAFII250) (TBP-associated factor 250 kDa) (P250) (Cell cycle gene 1
DE protein).
GN TAF2A OR CCG1 OR BAZR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=larineal carcinoma;
RX MEDLINE-91246200; PubMed-2038334;
RA Sekiguchi T., Nohiro Y., Nakamura Y., Hisamoto N., Nishimoto T.;
RT "The human CCG1 gene, essential for progression of the G1 phase,
RT encodes a 210-kilodalton nuclear DNA-binding protein.";
RL Mol. Cell. Biol. 11:3317-3325(1991).
RN [2]
RP PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE-89005056; PubMed-3169001;
RA Sekiguchi T., Miyata T., Nishimoto T.;
RT "Molecular cloning of the cDNA of human X chromosomal gene (CCG1)
RT which complements the temperature-sensitive G1 mutants, tsBN462 and
RT ts13, of the BHK cell line.";
RL EMBO J. 7:1683-1687(1988).
RN [3]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE-93196704; PubMed-7680771;
RA Rupert S., Wang E.H., Tjian R.;
RT "Cloning and expression of human TAFII250: a TBP-associated factor
RT implicated in cell-cycle regulation.";
RL Nature 362:175-179(1993).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE-93196705; PubMed-8450888;
RA Hisatake K., Hasegawa S., Takada R., Nakatani Y., Horikoshi M.,
RA Roeder R.G.;
RT "The p250 subunit of native TAFII box-binding factor TFIIID is the
RT cell-cycle regulatory protein CCG1.";
RL Nature 362:179-181(1993).
CC -1- FUNCTION: MAY PLAY AN ESSENTIAL ROLE IN TFIIID ASSEMBLY BY
CC INTERACTING WITH BOTH TBP AND OTHER TAF, AS WELL AS SERVING TO
CC LINK THE CONTROL OF TRANSCRIPTION TO THE CELL CYCLE. ESSENTIAL
CC FOR PROGRESSION OF THE G1 PHASE OF THE CELL CYCLE. POSSESSES
CC DNA-BINDING ACTIVITY.
CC -1- SUBUNIT: TF2D IS COMPOSED OF TBP AND A VARIETY OF TBP-ASSOCIATED
CC FACTORS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- PTM: PHOSPHORYLATED BY CASEIN KINASE II IN VITRO.
CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
CC -1- SIMILARITY: TO DROSOPHILA TAFII-230. SOME TO S.POMBE TAFII-111
CC AND TO S.CEREVISIAE TAF145.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90359; BAA14374.1; -
DR EMBL; X07024; CAA30073.1; ALT_SEQ.
DR PIR; S03005; S03005.
DR PIR; A40262; A40262.
DR TRANSFAC; T02206; -
DR MIM; 313650; -
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 2.

```

DR PROSITE: PSS0014; BROMODOMAIN_2; 2.
 KW Promotorin; Nuclear protein; DNA-binding; Cell cycle; Repeat;
 KM Transcription regulation; Phosphorylation.
 FT DNA_BIND 157 165
 FT DOMAIN 1195 1273 HMG BOX (POTENTIAL).
 FT DOMAIN 1351 1358 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 1397 1467 BROMODOMAIN 1.
 FT DOMAIN 1520 1590 BROMODOMAIN 2.
 FT DOMAIN 1627 1872 ASP/GLU-RICH (ACIDIC TAIL).
 SQ SEQUENCE 1872 AA; 212676 MW; 93BE3D181A72ABEB CRC64;

Query Match 17.0%; Score 145.5; DB 1; Length 1872;
 Best Local Similarity 26.0%; Pred. No. 0.00038;
 Matches 45; Conservative 33; Mismatches 72; Indels 23; Gaps 4;

QY 6 FSEFVTEIAPGYSMIKHPMDEFTMEKIKNDQSIIEELKDNFKIMCTNAMYKRPET 65
 DB 1528 FHHPVNNKKFVVDYKIVNPNDELTKRNISKHKYQSRSLDVLNLSVKNPES 1587
 QY 66 IYVAAKKLHSGMKILSQ--ERIQSLK-----QSIDFMADLQTRKQK 107
 DB 1588 QYTKAEIYVWCYQTLTEYDEHLTQLEKDICTAKEALEAEALIESLDPMTPGYTPQPP 1647
 QY 108 DGTDTSSG---GEGCGCMQRRERDS--GDAFAHAFKSPSKNKKKDMLEDK 155
 DB 1648 DLYDTNLSLSKRDASVAFQDESMSVLDIPSPATPEKOVYTOGEGEDGDLADEE 1700

RESULT 7
 SN24_HUMAN STANDARD; PRT; 1647 AA.
 AC P5133;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Possible global transcription activator SNF2L4 (SNF2-beta) (BRG-1 protein) (Mitotic growth and transcription activator) (Brahma protein homolog 1).
 GN SMARCA4 OR SNF2L4 OR BRG1 OR SNF2B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94050144; PubMed=8232556;
 RA Khavari P.A., Peterson C.L., Tamkun J.W., Mendel D.B., Crabtree G.R.;
 RT "BRG1 contains a conserved domain of the SM12/SNF2 family necessary for normal mitotic growth and transcription.";
 RL Nature 366:170-174(1993).
 RN [2]
 RP REVISIONS.
 RA Khavari P.A., Peterson C.L., Tamkun J.W., Mendel D.B., Crabtree G.R.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94268902; PubMed=820805;
 RA Chiba H., Muramatsu M., Nomoto A., Kato H.;
 RT "Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and Drosophila brama are transcriptional coactivators cooperating with the estrogen receptor and the retinoic acid receptor.";
 RL Nucleic Acids Res. 22:1815-1820(1994).
 RN [4]
 RP SEQUENCE OF 814-1474 FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V., Burkhart-Schultz K., Gordon L., Dias J., Brower A., Stillwagen S., Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J., Dangnan L., Ertler A., Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T., Trankhelm M., Amico-Keller G., Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B., Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S., Carrano A.V.;

RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: TRANSCRIPTIONAL COACTIVATOR COOPERATING WITH NUCLEAR HORMONE RECEPTORS TO POTENTIATE TRANSCRIPTIONAL ACTIVATION.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D29175; AAB40977.1; -;
 DR EMBL; D26156; BAA05143.1; -;
 DR EMBL; AC006127; AAC97987.1; -;
 DR MIM; 603254; -;
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR000330; SNF2_N.
 DR Pfam; PR00439; bromodomain_1.
 DR Pfam; PF00271; helicase_C_1.
 DR Pfam; PF00176; SNF2_N_1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELIC_C; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; 1.
 DR PROSITE; PSS0014; BROMODOMAIN_2; 1.
 KW Transcription regulation; Nuclear protein; Activator; Bromodomain;
 KM ATP-binding; Helicase.
 FT DOMAIN 578 588 POLY-LYS.
 FT DOMAIN 663 672 POLY-GLU.
 FT NP_BIND 779 786 ATP (POTENTIAL).
 FT SITE 881 884 DEGR BOX.
 FT DOMAIN 1360 1364 POLY-GLU.
 FT DOMAIN 1477 1547 BROMODOMAIN.
 FT DOMAIN 1571 1584 POLY-GLU.
 SQ SEQUENCE 1647 AA; 184585 MW; 7B785E7953277F1D CRC64;

Query Match 16.8%; Score 144; DB 1; Length 1647;
 Best Local Similarity 27.1%; Pred. No. 0.00044;
 Matches 39; Conservative 31; Mismatches 44; Indels 30; Gaps 5;

QY 16 PGYMIIRKHPMDEFTMEKIKNDQSIIEELKDNFKIMCTNAMYKRPETIYKAAKKL 75
 DB 1495 PEYELIRKPPVDFKIKIRIRNHRSLNDEKVMILCONAQTFNDEGSLYEDS---- 1550
 QY 76 HSGMKIISOERIOSKSIDPMADLQTRKOKDGTDSQSGEDGCMQRRERDSGDAEAH 135
 DB 1551 ----IYQSYFTYVRKIE-----KDD-----SEGEES-----EEEGEGERGS 1586
 QY 136 AFKSPSKENKKK--DKDMLDEDFK 157
 DB 1587 ESESRSVVKIKIKGRKEKAODRLK 1610

RESULT 8
 BRD3_HUMAN STANDARD; PRT; 726 AA.
 ID BRD3_HUMAN
 AC Q15059; Q92645;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Bromodomain-containing protein 3 (RING3-like protein).
 GN BRD3 OR RING3L OR KIA00043.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

OX NCBI_TaxID=9606;
RN
RP SEQUENCE FROM N.A..
RC TISSUE=BONE marrow;
RX MEDLINE=96051398; PubMed=7584044;
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA Seki N., Kawabayashi Y., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. II.
RT The coding sequences of 40 new genes (K1AA0041-K1AA0080) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 1:223-229(1994).
RN
RP SEQUENCE OF 363-726 FROM N.A.
RX MEDLINE=98038990; PubMed=9373153;
RA Thorpe K.L., Gorman P., Thomas C., Sheer D., Trowsdale J., Beck S.;
RT "Chromosomal localization, gene structure and transcription pattern of
RT the ORX gene, a homologue of the MHC-linked RING3 gene.";
RL Gene 200:177-183(1997).
CC -1 SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1 TISSUE SPECIFICITY: UBIQUITOUS.
CC -1 SIMILARITY: CONTAINS 2 BROMODOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D26362; BAA05393.1; -.
DR EMBL; Z81330; CAB03630.1; -.
DR MIM; 601541; -.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 2.
DR PROSITE; PS50014; BROMODOMAIN_2; 2.
KW Bromodomain; Repeat; Nuclear protein.
FT DOMAIN 56 115 BROMODOMAIN 1.
FT FT 326 398 BROMODOMAIN 2.
FT DOMAIN 487 555 LYS-RICH.
FT DOMAIN 676 725 SER-RICH.
FT CONFLICT 465 466 EL->DY (IN REF. 2).
SO SEQUENCE 726 AA; 79541 MW; 64F526FC3C1033AA CRC64;

Query Match 16.6%; Score 142; DB 1; Length 726;
Best Local Similarity 32.7%; Pred. No. 0.00024;
Matches 33; Conservative 17; Mismatches 47; Indels 4; Gaps 1;

OY 16 PEYSMIKHPMDEFSTWKEIKIKNDYQSIIEELKDNFKLMCTNMIYKPEPTIYKAKKIL 75
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DBb 71 PDYHIILNPMDMGRIKRLLENYYWASASECQMDNTMTFTNYIKKPTDIDIVLAQAL- 129
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 76 HSGMKILSGERISLKQSIDEMADLQIKRKOKDGTDSQSG 116
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 130 ---EKIFLQKVAQMPQEEVEYELPPAPKGRKRPAAQAQSAQ 167
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
BRD2_HUMAN ID BRD2_HUMAN STANDARD; PRT; 801 AA.
AC P25440;
DT 01-MAY-1992 (Rel. 22, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bromodomain-containing protein 2 (RING3 protein).
DE BRD2 OR RING3 OR KIAA9001.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

Ox	NCDL_TaxID=96606;
Rn	[1]
Rp	SEQUENCE FROM N.A.
Rc	TISSUE-T-cell;
Rx	MEDLINE=92329974; PubMed=1352711;
Ra	Beck S., Hanson I., Kelly A., Pappin D.J.C., Trowsdale J.;
Rt	"A homologue of the Drosophila female sterile homeotic (<i>fsh</i>) gene in
Rl	the class II region of the human MHC.";
Rf	DNA Seq. 2:203-210(1992).
Rn	[2]
Rm	REVISIONS TO N-TERMINUS.
Rx	MEDLINE=96376536; PubMed=8781126;
Ra	Thorpe K.L., Abdulla S., Kaufman J., Trowsdale J., Beck S.;
Rt	"Phylogeny and structure of the RING3 gene.";
Rl	Immunogenetics 44:391-396(1996).
Rn	[3]
Rp	SEQUENCE FROM N.A.
Rc	TISSUE=Bone marrow;
Ra	Nonuma N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,
Rt	Sato S., Nagase T., Seki T., Ishikawa K.-I., Tabata S.;
Rl	Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.
Rc	-1- SUBCELLULAR LOCATION: Nuclear (potential).
Cc	-1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
Cc	-1- SIMILARITY: CONTAINS 1 ET DOMAIN.
Cc	-----
Cc	This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc	between the Swiss Institute of Bioinformatics and the EMBL outstation -
Cc	the European Bioinformatics Institute. There are no restrictions on its
Cc	use by non-profit institutions as long as its content is in no way
Cc	modified and this statement is not removed. Usage by and for commercial
Cc	entities requires a license agreement (see http://www.isb-sdb.ch/announce/
Cc	or send an email to license@sdb-slb.ch).
Cc	-----
Dr	EMBL; X62083; CAA43396.1; -
Dr	EMBL; M80613; AAA68890.1; ALT_INTX.
Dr	EMBL; D42040; BAA07641.1; -
Dr	PIR; S18860; S18860.
Dr	PIR; S40781; S40781.
Dr	MIM; 601540; -
Dr	InterPro: IPR001487; Bromodomain.
Dr	Pfam: PF00439; bromodomain; 2.
Dr	PRINTS: PR00503; BROMODOMAIN.
Dr	SMART: SMO0297; BROMO; 2.
Dr	PROSITE: PS00633; BROMODOMAIN_1; 2.
Dr	PROSITE: PS50014; BROMODOMAIN_2; 2.
Rw	Bromodomain; Repeat; Nuclear protein.
Ft	DOMAIN 91 163 BROMODOMAIN 1.
Ft	DOMAIN 364 436 GLU/SER-RICH.
Ft	DOMAIN 476 515 SER/RICH.
Ft	DOMAIN 544 566 ARG/LYS-RICH (HIGHLY BASIC).
Ft	DOMAIN 775 801 SER-RICH.
Ft	DOMAIN 555 559 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
Ft	DOMAIN 638 801 ET DOMAIN.
Ft	DOMAIN 61 64 POLY-PRO.
Ft	DOMAIN 492 506 POLY-GLU.
Ft	DOMAIN 551 559 POLY-LYS.
Ft	DOMAIN 634 638 POLY-GLU.
Ft	DOMAIN 775 793 POLY-SER.
So	SEQUENCE 801 AA; 88060 MW; 9A075EB13507D8E CRC64;

Query Match 16.6%; Score 142; DB 1; Length 801;
Best Local Similarity 25.3%; Pred. No. 0.00027;
Matches 44; Conservative 20; Mismatches 66; Indels 42; Gaps 2;

Oy	18	YSMIKHQPDSTMEKIKNDYOSIEELKONFKIMCTNATMYNKPEITIIYAARKL---	74
Db	386	YHDIKHEMDISTVYRKKNRDYRPAOEAFADVRILFMSNCYKYNPDPVVYVAAMKLADV	445
Oy	75	---LHGSMIIISGERIQSLKQSIDFMADIQTKFKQKDGDGTSGSGAGDGCGW	122
Db	446	FEERYAKMPDEPLEGPPLPVSTAMPPLGALKSESSSESSSESSSESSSEEEEEEDEDEDEEE	505

OY 123 QREDSGDAAEHAHAF-----KSPSKENKKKK 149
 Db 506 ESESDSEERAHRLAEOLRAVHEOLALSGPSKPKRKKKKKKKK 559
 RESULT 10
 FSH_DROME STANDARD: PRT: 2038 AA.
 ID FSH_DROME
 AC P13709; P13710; 13, Created
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Female sterile homeotic protein (Fragile-chorion membrane protein).
 GN FSH1H OR FSH.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=69276730; PubMed=2567251;
 RA Haynes S.R., Mozer B.A., Bhatia-Dey N., David I.B.;
 RT "The Drosophila fsh locus, a maternal effect homeotic gene, encodes
 RT apparent membrane proteins.";
 RL Dev. Biol. 134:246-257(1989).
 CC -1- FUNCTION: REQUIRED MATERIALLY FOR PROPER EXPRESSION OF OTHER
 CC HOMEOTIC GENES INVOLVED IN PATTERN FORMATION, SUCH AS UBX.
 CC -1- SIMILARITY: HIGH, TO HUMAN RING3 PROTEIN.
 CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 ET DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M23221; AAA28540.1; -;
 DR EMBL: M23222; AAA28541.1; ALT_TERM.
 DR EMBL: M15762; AAA70424.1; -;
 DR EMBL: M15763; AAA70423.1; -;
 DR EMBL: M15764; AAA70422.1; -;
 DR PIR: A43742; A43742.
 DR HSP: P04002; IYFA.
 DR FlyBase: FBgn0004656; fs(1)h.
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; bromodomain; 2.
 DR PRINTS: PRO0503; BROMODOMAIN.
 DR SMART: SM00297; BROMO: 2.
 DR PROSITE: PS00633; BROMODOMAIN_1; 2.
 DR PROSITE: PS50014; BROMODOMAIN_2; 2.
 KW Developmental Protein; Bromodomain; Transmembrane; Repeat.
 FT DOMAIN 51 123 BROMODOMAIN 1.
 FT DOMAIN 495 567 BROMODOMAIN 2.
 FT DOMAIN 945 1106 ET DOMAIN.
 FT TRANSMEM 330 350 POTENTIAL.
 FT TRANSMEM 451 471 POTENTIAL.
 FT TRANSMEM 750 770 POTENTIAL.
 FT TRANSMEM 790 810 POTENTIAL.
 FT TRANSMEM 816 830 POTENTIAL.
 FT TRANSMEM 874 894 POTENTIAL.
 FT TRANSMEM 1731 1751 POTENTIAL.
 FT TRANSMEM 1939 1959 POTENTIAL.
 FT VARIANT 909 909 G -> A.
 FT VARIANT 1022 1022 H -> RPY.
 SQ SEQUENCE 2038 AA; 205332 MW; 849E0706D50A0098 CRC64;

Query Match 15.88; Score 135; DB 1; Length 2038;
 Best Local Similarity 28.44; Pred. No. 0.0031;

Matches 42; Conservative 22; Mismatches 70; Indels 14; Gaps 5;
 OY 6 FSPFYDFI-----APGSMIIRHMPDFSTMKRKNNDYSTEELKDNFKLCTNAMI 59
 Db 55 FSWPFOQPVDAKKLNLDPYHKIIIPQMDMGITIKRLENNYWSAKETIQDNTFNHCYV 114
 OY 60 YNKEETIYKRAKKLLSGMKILSOERIQSL-KQSIDFMADLQTRKQDGTDSGGED 118
 Db 115 YNKPGEDEVVVAQQL----EKVPLQ-KIESMPKELELEPYTAKGCKKORAPATPKSSS 169
 OY 119 GGCWQREDSGDAAEHAHAFKSPSKENKK 146
 Db 170 GGA--GASTGSGTSSAAVTSQPGSGSTK 195
 RESULT 11
 GCNS_YEAST STANDARD: PRT: 439 AA.
 ID GCNS_YEAST
 AC 003330;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Transcriptional activator GCNS.
 GN GCNS OR ADA4 OR YGR252M.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_Taxid=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93011009; PubMed=1396595;
 RA Georgakopoulos T., Thireos G.;
 RT "Two distinct yeast transcriptional activators require the function
 RT of the GCNS protein to promote normal levels of transcription.";
 RL EMO J. 11:415-4152(1992).
 [2]
 RP SEQUENCE OF 1-170 FROM N.A.
 RC STRAIN=S288C / FY1679;
 RX MEDLINE=97279234; PubMed=9133742;
 RA Petrol F., Garlignani G., Pavanello A., Guerrelro P., Azevedo D.,
 RA Rodrigues-Pousada C., Melchiorretto P., Panzeri L.,
 RA Agostoni Carbone M.L.;
 RT "Analysis of a 17.9 kb region from Saccharomyces cerevisiae
 RT chromosome VII reveals the presence of eight open reading frames,
 RT including BRP1 (TFIIB/0) and GCNS genes.";
 RL Yeast 13:373-377(1997).
 RN [3]
 RP SEQUENCE OF 170-439 FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=97279233; PubMed=9133741;
 RA Mazzoni C., Ruzzi M., Rinaldi T., Solinas F., Montedove F.,
 RA Frontali L.;
 RT "Sequence analysis of a 10.5 kb DNA fragment from the yeast
 RT chromosome VII reveals the presence of three new open reading frames
 RT and of a tRNAIle gene.";
 RL Yeast 13:369-372(1997).
 RN [4]
 RP ASSOCIATION WITH ADA2.
 RX MEDLINE=95045371; PubMed=7957049;
 RA Marcus G.A., Silverman N., Berger S.L., Horluchi J., Guarente L.;
 RT "Functional similarity and physical association between GCNS and
 RT ADA2: putative transcriptional adaptors.";
 RL EMO J. 13:4807-4815(1994).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 99-262.
 RX MEDLINE=99362688; PubMed=10430873;
 RA Trivelp R.C., Rojas J.R., Steiner D.E., Venkataramani R.N., Wang L.,
 RA Zhou J., Allis C.D., Berger S.L., Marmorstein R.;
 RT "Crystal structure and mechanism of histone acetylation of the yeast
 RT GCNS transcriptional activator.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:8931-8936(1999).
 CC -1- FUNCTION: GENERAL TRANSCRIPTIONAL ACTIVATOR OPERATING IN CONCERT
 CC WITH CERTAIN OTHER DNA-BINDING TRANSCRIPTIONAL ACTIVATORS SUCH AS

CC GCN4 OR HAP2/3/4. FUNCTION AS AN HISTONE ACETYLTRANSFERASE (HAT)
CC TO PROMOTE TRANSCRIPTIONAL ACTIVATION. HAS A STRONG PREFERENCE FOR
CC LYSINE 14 OF H3 AND A SOMEWHAT LOWER PREFERENCE FOR LYSINES 8 AND
CC 16 OF HISTONE H4.
CC -1- SUBUNIT: HETERO-DIMER WITH ADA2. PART OF THE ADA/GCN5 COMPLEX THAT
CC CONSISTS OF HPI1/ADA1, ADA2, ADA3, SPT20/ADA5 AND GCN5.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X68628; CAA48602.1; -
DR EMBL: Z73037; CAA97281.1; -
DR EMBL: X99228; CAA67614.1; -
DR PIR: S28051; S28051.
DR PDB: 1YGH: 02-AUG-99.
DR TRANSFAC: T02145; -
DR SGD: S0003484; GCN5.
DR InterPro: IPR00182; Acetyltransf_GCN5.
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00583; Acetyltransf_1.
DR Pfam: PF00439; bromodomain; 1.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 1.
DR PROSITE: PS00633; BROMODOMAIN_1; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
DR Transcription regulation; DNA-binding; Activator; Trans-acting factor;
KW 3D-structure.
FT ACT_SITE 173 173 GENERAL BASE.
FT DOMAIN 344 414 BROMODOMAIN.
SQ SEQUENCE 439 AA; 51069 MW; 3200730DDC7EE70D CRC64;

Query Match 15.2%; Score 130; DB 1; Length 439;
Best Local Similarity 40.6%; Pred. No. 0.0013;
Matches 28; Conservative 11; Mismatches 30; Indels 0; Gaps 0;

QY 6 FSPFVDTAFGYSMITIKHPDMFSTMKERIKNNQVSIIELDKFNKCTNMTYKPPET 65
DB 352 FQAPVKEKEVPDYDFIKPEPMDLSTWEIKLESNKYQKMEFDYDAIVNNGCMYNGENT 411
QY 66 IYYKAKKL 74
DB 412 STYKIANRL 420

RESULT 12
CAP_MOUSE STANDARD; PRT; 2441 AA.
AC P45481;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE CREB-binding protein.
GN CREBBP OR CBP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=94019866; PubMed=8413673;
RA Christia J.C., Kwok R.P.S., Lamb N., Haglwaara M., Montminy M.R.,
RA Goodman R.H.;
RT "Phosphorylated CREB binds specifically to the nuclear protein CBP";
RL Nature 365:855-859(1993).

CC -1- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS
CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
CC CAMP-RESPONSIVE GENES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -1- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S66385; AAB28651.1; -
DR TRANSFAC: T01318; -
DR MGD: MGI:1098280; Crebpb.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR003101; KIX.
DR InterPro: IPR000197; TAF_finger.
DR InterPro: IPR000433; ZnF_ZZ.
DR Pfam: PF00439; bromodomain; 1.
DR Pfam: PF02172; KIX; 1.
DR Pfam: PF02135; ZF-TAZ; 2.
DR Pfam: PF00569; Z2; 1.
DR Pfam: PF00503; BROMODOMAIN.
DR PRINTS: SM00297; BROMO; 1.
DR SMART: SM00297; BROMO; 1.
DR PROSITE: PS00633; BROMODOMAIN_1; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
DR PROSITE: PS01357; ZF_ZZ_1; 1.
DR PROSITE: PS50135; ZF_ZZ_2; 1.
KW Transcription regulation; Nuclear protein; Activator; Bromodomain;
KW Zinc-finger.
FT DOMAIN 1104 1176 BROMODOMAIN.
FT ZN_FING 1702 1745 Z2-TYPE.
FT DOMAIN 1062 1065 POLY-GLU.
FT DOMAIN 1556 1563 POLY-GLU.
FT DOMAIN 1944 1949 POLY-PRO.
FT DOMAIN 1968 1971 POLY-GLN.
FT DOMAIN 2082 2086 POLY-GLN.
FT DOMAIN 2200 2216 POLY-GLN.
FT DOMAIN 2296 2299 POLY-GLN.
SQ SEQUENCE 2441 AA; 265474 MW; 0AB3028C3112F419 CRC64;

Query Match 15.1%; Score 129; DB 1; Length 2441;
Best Local Similarity 27.0%; Pred. No. 0.012; Mismatches 69; Indels 44; Gaps 9;
Matches 51; Conservative 25;

QY 1 DSAF-FSPFVDTAF--APGYSMITIKHPDMFSTMKERIKNNQVSIIELDKFNKCTNNA 57
DB 1106 DPESLPFRQPVDPQLGIDYDIYKNPMDLSTIKKKLDTGQYQEWQYVVDVRLMFNNA 1165
QY 58 MIYNKPEIYYKAAKLLISGKILISQERISQSIDMADLQTRKKDGTDT----- 112
DB 1166 WLYNRKTSYVYKFSKLT---AEVEFOE-IDPYMOSLIGCCG---RKYEFSPTLCCYG 1216
QY 113 -----SOSGEDGCGW---QREEDSGDAEAAFAKSPSENNKKDKMDLE 153
DB 1217 KOLCTIPRDAAYSYONRHHFGCKTEIQGENVILGD-----DPSQPTTISKDPE 1269
QY 154 DKFSNNLE 162
DB 1270 KK-KNDTLD 1277

RESULT 13
BRD4_HUMAN STANDARD; PRT; 1362 AA.
ID BRD4_HUMAN
AC 060885; Q96PD3;

DT 16-OCT-2001 (Rel. 40, Last Created)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Bromodomain-containing protein 4 (HUNK1 protein).
 GN BRD4 OR HUNK1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NC NCB1_TaxID=9606;
 RN NCB1
 RP SEQUENCE FROM N.A.
 RA French C.A., Fletcher J.A.;
 RT "Human BRD4 protein."
 RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-722 FROM N.A.
 RC TISSUE=Placenta;
 RA Weber B.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF386649; AAL26987.1; -
 DR EMBL: Y12059; CA72780.1; -
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; Bromodomain; 2.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00297; BROMO; 2.
 DR PROSITE: PS00633; BROMODOMAIN_1; 1.
 DR PROSITE: PS00014; BROMODOMAIN_2; 2.
 KM Bromodomain; Repeat; Nuclear protein.
 FT DOMAIN 1
 FT 368 440 BROMODOMAIN 1.
 FT 535 594 BROMODOMAIN 2.
 FT 692 717 LYS-RICH.
 FT 703 714 SER-RICH.
 FT 738 743 POLY-SER.
 FT 757 761 POLY-HIS.
 FT 764 770 POLY-PRO.
 FT 771 775 POLY-GLN.
 FT 776 783 POLY-PRO.
 FT 954 964 POLY-PRO.
 FT 974 986 POLY-PRO.
 FT 1011 1014 POLY-PRO.
 FT 1028 1033 POLY-PRO.
 FT 1283 1300 POLY-GLN.
 FT 1301 1308 POLY-ALA.
 FT 1335 1338 POLY-ARG.
 FT 720 721 EM -> GP (IN REF. 2).
 SO SEQUENCE 1362 AA; 152219 MW; D52EFELCF960907 CMC64;

Query Match 14.8%; Score 126.5; DB 1; Length 1362;
 Best Local Similarity 21.2%; Pred. No. 0.0094;
 Matches 43; Conservative 33; Mismatches 52; Indels 75; Gaps 5;

QY 18 YSMITKHPDFTMKKEIKKNDYQSEIEELKDNFKLMCTNMTINKPETIYKAKKU--- 74
 DB 390 YCDIIRHPDMSTIKSKLEAREYRDQEFADAVRLMFMSNCKYKPNPDHEVVAARKLQDV 449
 QY 75 ----LHSGMKILTS----- 83
 DB 450 FEMRAKMPDEPEEPVAVSPAPPTKVVAVAPSSSDSSSDSSSDSSDSTDSSEERRAQ 509
 QY 84 -----QRIOSLQKSIDFMADLQTRKQKDGDTDSGEGCGQWREREDSGDAEAHAFK 138

DB 510 RLAELEQULKAHVHQAALASQPOONKPKKREKD-----KKEKK---KEKHRRK 554
 QY 139 SPSEKRNK-RKDKMLDEDEKSN 160
 DB 555 EEVEENKSKAKPEPPPKTKKN 577
 RESULT 14
 ID BRM_DROME STANDARD; PRT; 1638 AA.
 AC P25439; O9VIM5; O9VIM6;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Homeotic gene regulator (Brahma protein).
 GN BRM OR CG5942 OR CG18438.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCB1_TaxID=7227;
 RN NCB1
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RX MEDLINE=92154670; PubMed=1346755;
 RA Tamkun J.W., Deuring R., Scott M.P., Kissinger M., Pattatucci A.M.,
 RA Kautman T.C., Kennison J.A.;
 RT "Brahma: a regulator of Drosophila homeotic genes structurally
 RT related to the yeast transcriptional activator SNF2/SWI2."
 RL Cell 68:561-572(1992).
 RN [2]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Abrial J.F., Adysvany A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Teclor C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: MAY ACT AS COACTIVATOR, ASSISTING ONE OR MORE DEDICATED

CC TRANSCRIPTIONAL ACTIVATORS OF ANT-C AND BX-C HOMEOTIC GENES. CAN
 CC CONTACT THE REPRESSIVE EFFECT OF POLYCOMB PROTEIN.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- DEVELOPMENTAL STAGE: HIGHEST EXPRESSION IN UNFERTILIZED EGGS AND
 CC EARLY EMBRYOS.
 CC -1- MISCELLANEOUS: 'BRAHMA' MEANS 'FATE' IN INDIA.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M85049; AAA19661.1; -
 CC EMBL; AE003529; AAF49557.1; -
 CC EMBL; AE003529; AAF49558.2; -
 CC PIR; A42091; A42091
 CC Flybase; FBgn0000212; btm.
 CC InterPro; IPR001487; Bromodomain.
 CC InterPro; IPR001410; DEAD.
 CC InterPro; IPR001650; Helicase_C.
 CC InterPro; IPR000330; SNF2_N.
 CC Pfam; PF00439; bromodomain; 1.
 CC Pfam; PF00271; helicase_C; 1.
 CC Pfam; PF00176; SNF2_N; 1.
 CC PRINTS; PR00503; BROMODOMAIN.
 CC SMART; SM00287; BROMO; 1.
 CC SMART; SM00490; HELIC_C; 1.
 CC PROSITE; PS00633; BROMODOMAIN_1; 1.
 CC PROSITE; PS00633; BROMODOMAIN_2; 1.
 CC KMW Actinuator; Bromodomain; Transcription regulation; Nuclear protein;
 CC KMW Actinuator; Bromodomain; Helicase; ATP-binding; Alternative splicing.
 CC FT DOMAIN 201 390
 CC FT NP BIND 798 805 ATP (POTENTIAL).
 CC FT SITE 900 903 DECH BOX.
 CC FT DOMAIN 1385 1392 ASP/GLU-RICH (ACIDIC).
 CC FT DOMAIN 1394 1404 ARG/LYS-RICH (BASIC).
 CC FT DOMAIN 1405 1410 ASP/GLU-RICH (ACIDIC).
 CC FT DOMAIN 1415 1432 ARG/LYS-RICH (BASIC).
 CC FT DOMAIN 1443 1513 BROMODOMAIN.
 CC FT DOMAIN 1631 1638 POLY-ASP.
 CC FT VARSPIC 121 121 G->A (IN SHORT ISOFORM).
 CC FT VARSPIC 122 222 MISSING (IN SHORT ISOFORM).
 CC FT CONFLICT 687 687 D->Y (IN REF. 1).
 CC SQ SEQUENCE 1638 AA; 18508 MW; A4494B29FAF2EA2A CRC64;

Query Match 14.5%; Score 124; DB 1; Length 1638;
 Best Local Similarity 28.8%; Pred. No. 0.019;
 Matches 40; Conservative 21; Mismatches 60; Indels 18; Gaps 4;

ID CBP_HUMAN STANDARD; PRT; 2442 AA.
 AC 092793; Q16376; Q00147;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE CREB-binding protein.
 GN CREBBP OR CBP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NCBI_Taxid=9606;
 RX MEDLINE=97385172; PubMed=9238046;
 RX Schlobo O.M., Borrow J., Tomek R., Reshimi S., Harden A.,
 RA Schlegelberger B., Housman D., Doggett N.A., Rowley J.D.,
 RA Zelenik-Le N.J.;
 RT "MLL is fused to CBP, a histone acetyltransferase, in therapy-related
 RT acute myeloid leukemia with a t(11;16)(q23;p13.3).";
 RT Proc. Natl. Acad. Sci. U.S.A. 94:8732-8737(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=97321049; PubMed=9177780;
 RX Giles R.H., Petrij F., Dauwerse H.G., den Hollander A.I.,
 RA Lushnikova T., van Ommen G.J.B., Goodman R.H., Deaven L.L.,
 RA Doggett N.A., Peters D.J.M., Breuning M.H.;
 RT "Construction of a 1.2-Mb contig surrounding, and molecular analysis
 RT of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome
 RT 16p13.3.";
 RL Genomics 42:96-144(1997).
 RN [3]
 RP SEQUENCE OF 1-405 FROM N.A.
 RP MEDLINE=96376968; PubMed=8782817;
 RX Borrow J., Stanton V.P., Andersen J.M., Becher R., Behn F.G.,
 RA Chaganti R.S.K., Clavin C.I., Distcheche C., Dube I., Fritschauf A.M.,
 RA Housman D., Mittleman F., Volinia S., Watson A.E., Housman D.E.;
 RT "The translocation t(8;16)(p11;p13) of acute myeloid leukemia fuses
 RT a putative acetyltransferase to the CREB-binding protein.";
 RL Nat. Genet. 14:33-41(1996).
 CC -1- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
 CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS
 CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
 CC CAMP-RESPONSIVE GENES.
 CC -1- SUBUNIT: INTERACTS WITH SMAD1, SMAD2 AND SMAD3.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
 CC T(8;16)(p11;p13) INVOLVING CBP AND MOZ, AND T(11;16)(q23;p13.3)
 CC INVOLVING CBP AND MLL.
 CC -1- DISEASE: DEFECTS IN CREBBP ARE THE CAUSE OF RUBINSTEIN-TAYBI
 CC SYNDROME (RITS), A DISORDER CHARACTERIZED BY CRANIOFACIAL
 CC ABNORMALITIES, BROAD THUMBS, BROAD BIG TOES, MENTAL RETARDATION
 CC AND A PROPENSITY FOR DEVELOPMENT OF MALIGNANCIES.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U47741; AAC51770.1; -
 CC EMBL; U85962; AAC51331.1; -
 CC EMBL; U89354; AAC51339.1; -
 CC EMBL; U89355; AAC51340.1; -
 CC MIM; 600140; -
 CC InterPro; IPR001487; Bromodomain.
 CC InterPro; IPR003101; KIX.
 CC InterPro; IPR000197; TAZ_finger.
 CC InterPro; IPR000433; znf_zz.
 CC Pfam; PF00439; bromodomain; 1.

DR	Pfam: PF02172; KIX; 1.	
DR	Pfam: PF02135; ZF-TAZ; 2.	
DR	Pfam: PF00569; ZF; 1.	
DR	PRINTS: PR00503; BROMODOMAIN.	
DR	SMART; SM00297; BROMO; 1.	
DR	SMART; SM00291; ZNF_ZF; 1.	
DR	PROSITE: PS00633; BROMODOMAIN_1; 1.	
DR	PROSITE: PS00614; BROMODOMAIN_2; 1.	
DR	PROSITE: PS01357; ZF_ZF_1; 1.	
DR	PROSITE: PS01357; ZF_ZF_2; 1.	
KW	Transcription regulation; Nuclear protein; Activator; Bromodomain	
KW	Chromosomal translocation; Zinc-finger.	
FT	FT ZN_FING	1701 1744
FT	DOMAIN	363 430
FT	DOMAIN	452 683
FT	DOMAIN	1103 1175
FT	DOMAIN	1061 1064
FT	DOMAIN	1199 1487
FT	DOMAIN	1555 1562
FT	DOMAIN	1675 1849
FT	DOMAIN	1943 1948
FT	DOMAIN	1967 1970
FT	DOMAIN	2081 2085
FT	DOMAIN	2199 2216
FT	DOMAIN	2245 2248
FT	DOMAIN	2297 2300
FT	CONFLICT	1511 1513
FT	CONFLICT	1724 1725
FT	CONFLICT	1770 1770
FT	CONFLICT	1789 1789
FT	CONFLICT	1812 1812
SO	SEQUENCE	265336 AA; 420084619475F3D2 CRC64;

Query Match	14.4%	Score 123;	DB 1;	Length 2442;
Best Local Similarity	34.3%;	Pred. No. 0.036;		
Matches	34;	Conservative	16;	Mismatches 41;
			Indels	8;
			Gaps	4

QY 1 DPSAF--SSEPTDTDI--APQSMIIKHMPDQSFMSRKIKIANNQYQSIIEELKDNFKETLACMGNA 57
Db 1105 DPESLPAPRQPDQDLGIPYFDIVANNPMDLSIRKKLDTGQYQERWQYVDYVWLMFPNNA 1164
QY 58 MIYMKPETIYKAAKKLLHSGMKMLIISOERQSIQKSIDF 96
Db 1165 WLNIRKTSRYKFCPSKI---ALEFDE--IDPQMSLQ 1198

Search completed: July 11, 2002, 15:55:26
Job time: 349 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 15:56:16 ; Search time 49.31 Seconds

(without alignments)
571.855 Million cell updates/sec

Title: US-09-687-230-2_COPY_151_313

Perfect score: 855
Sequence: 1 DPSAFSFPVDTFTAPGYSM.....NKKKDKMLDEKSKNNLEIR 163

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	855	100.0	651	4 Q9NP11	Q9NP11 homo sapien
2	855	100.0	652	4 Q9UH59	Q9UH59 homo sapien
3	777	90.9	651	11 O88665	O88665 mus musculu
4	610	71.3	459	4 Q9BV48	Q9BV48 homo sapien
5	375	43.9	501	4 Q9H8M2	Q9H8M2 homo sapien
6	328	38.4	233	4 Q9H7R9	Q9H7R9 homo sapien
7	284.5	33.3	861	5 Q9VLX2	Q9VLX2 drosophila
8	200.5	23.5	636	5 Q17581	Q17581 caenorhabd1
9	174.5	20.4	952	10 Q9SNT9	Q9SNT9 oryza sativ
10	174	20.4	715	4 Q9V4O3	Q9V4O3 homo sapien
11	174	20.4	805	4 Q95692	Q95692 homo sapien
12	172.5	20.2	706	11 Q9SVV4	Q9SVV4 mus musculu
13	169.5	19.8	652	10 Q9LM88	Q9LM88 arabidopsis
14	166	19.4	145	10 Q9FIA2	Q9FIA2 arabidopsis
15	163.5	19.1	556	10 Q9SEF1	Q9SEF1 arabidopsis
16	160	18.7	1430	5 Q9VAJ4	Q9VAJ4 drosophila

17	156	18.2	473	5 Q9Y0F2	Q9Y0F2 toxoplasma
18	156	18.2	1169	5 Q9NTS0	Q9NTS0 toxoplasma
19	155.5	18.2	1568	13 Q90755	Q90755 gallus gall
20	151.5	17.7	757	4 O15355	O15355 homo sapien
21	149.5	17.5	418	5 O27198	O27198 tetrahymena
22	148.5	17.4	236	11 Q9D007	Q9D007 mus musculu
23	148.5	17.4	454	3 Q9U0K2	Q9U0K2 schizosach
24	148.5	17.4	1235	4 Q9H0E9	Q9H0E9 homo sapien
25	147	17.2	513	5 Q9VCG6	Q9VCG6 drosophila
26	147	17.2	1630	13 Q90753	Q90753 gallus gall
27	146.5	17.1	729	13 Q90971	Q90971 gallus gall
28	146.5	17.1	733	13 Q73897	Q73897 gallus gall
29	146	17.1	920	4 Q43178	Q43178 homo sapien
30	146	17.1	920	4 Q969M9	Q969M9 homo sapien
31	146	17.1	2065	5 Q97068	Q97068 drosophila
32	144	16.8	1022	11 Q63928	Q63928 mus sp. brg
33	144	16.8	1647	4 Q9HBD3	Q9HBD3 homo sapien
34	144	16.8	1654	5 Q9VC36	Q9VC36 drosophila
35	144	16.8	1679	4 Q9HBD4	Q9HBD4 homo sapien
36	142	16.6	726	11 Q9J125	Q9J125 mus musculu
37	142	16.6	754	4 Q15310	Q15310 homo sapien
38	142	16.6	801	4 Q969U4	Q969U4 homo sapien
39	141.5	16.5	3080	5 Q9VRY3	Q9VRY3 drosophila
40	140	16.4	374	5 Q76561	Q76561 caenorhabd1
41	139	16.3	1792	5 Q9YU19	Q9YU19 caenorhabd1
42	138.5	16.2	798	11 Q54795	Q54795 mus musculu
43	138.5	16.2	798	11 Q88411	Q88411 mus musculu
44	135.5	15.8	1865	11 Q60544	Q60544 mesocricetu
45	135	15.8	1937	5 Q9W3L3	Q9W3L3 drosophila

ALIGNMENTS

RESULT 1
ID Q9NP11 PRELIMINARY; PRT; 651 AA.
AC Q9NP11;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE BROMODOMAIN CONTAINING PROTEIN (NAG4).
GN BR465.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FETAL BRAIN.
RA Kzyshkowska J.G., Dobner T.G.;
RT "Cloning of a human bromodomain containing protein."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yu Y., Li G.Y.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; A2721881; CAB72445.1; -
DR EMBL; AF152604; AAF75126.1; -
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR PROSITE; PSS0014; BROMODOMAIN_2; 1.
SQ SEQUENCE 651 AA; 74138 MW; 29B7947644C215E7 CRC64;

Query Match 100.0%; Score 855; DB 4; Length 651;
Best Local Similarity 100.0%; Pred. No. 2.8e-63;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DPSAFSFPVDTFTAPGYSMLIKHPMDFSTMKERIKNNYQSIELDKNFKMTNMIY 60
|||||

```

Db 151 DSAFFSPVTDPIAGYSMIKHHPDFTMKKIKNNNDYOSIEELKDNFKLMCTNAMY 210
OY 61 NKPEITYYKAARKLLHSGMKILISOERIOSLKOSIDPMADLOKTRKOKDGTDSOGSDGG 120
    |||||||
Db 211 NKPEITYYKAARKLLHSGMKILISOERIOSLKOSIDPMADLOKTRKOKDGTDSOGSDGG 270
OY 121 CMOEREDSGDAEAHAFKSPSKENKKDKMDLEDFKSNLNR 163
    |||||||
Db 271 CMOEREDSGDAEAHAFKSPSKENKKDKMDLEDFKSNLNR 313

RESULT 2
O9UH59 PRELIMINARY: PRT; 652 AA.
AC 09UH59;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE BROMODOMAIN PROTEIN CELTIX1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Stahl A., Enserink J., Stein J.L., Stein G.S., van Wijnen A.J.;
RT "Molecular analysis of the human bromodomain protein Celtix-1.";
RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF213969; AAF19526.1; -.
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; bromodomain; 1.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
SQ SEQUENCE 652 AA; 74084 MW; 70F2B654B2618529 CRC64;

Query Match 100.0%; Score 855; DB 4; Length 652;
Best Local Similarity 100.0%; Pred. No. 2.8e-63;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DSAFFSPVTDPIAGYSMIKHHPDFTMKKIKNNNDYOSIEELKDNFKLMCTNAMY 60
    |||||||
Db 151 DSAFFSPVTDPIAGYSMIKHHPDFTMKKIKNNNDYOSIEELKDNFKLMCTNAMY 210
OY 61 NKPEITYYKAARKLLHSGMKILISOERIOSLKOSIDPMADLOKTRKOKDGTDSOGSDGG 120
    |||||||
Db 211 NKPEITYYKAARKLLHSGMKILISOERIOSLKOSIDPMADLOKTRKOKDGTDSOGSDGG 270
OY 121 CMOEREDSGDAEAHAFKSPSKENKKDKMDLEDFKSNLNR 163
    |||||||
Db 271 CMOEREDSGDAEAHAFKSPSKENKKDKMDLEDFKSNLNR 313

RESULT 3
O88665 PRELIMINARY: PRT; 651 AA.
AC 088665;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE BROMODOMAIN-CONTAINING PROTEIN BP75.
CN BRD7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-BALB/C; TISUE-BRAIN;
RC MEDLINE=99456832; PubMed=10526152;
RA Cuppen E., van Ham M., Peeters B., Wieringa B., Hendriks W.;
RT "Identification and molecular characterization of BP75, a novel

```

```

RT bromodomain-containing protein.";
RL FEBS Lett. 459:291-298(1999).
DR EMBL; AF084259; AAC33302.1; -.
DR MGI; MGI:1349766; Brd7.
DR InterPro: IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 1.
DR SMART; SM00297; BROMO; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
SQ SEQUENCE 651 AA; 74000 MW; 5D34B4F14FD51350 CRC64;

Query Match 90.9%; Score 777; DB 11; Length 651;
Best Local Similarity 89.6%; Pred. No. 8.7e-57;
Matches 146; Conservative 10; Mismatches 7; Indels 0; Gaps 0;

OY 1 DSAFFSPVTDPIAGYSMIKHHPDFTMKKIKNNNDYOSIEELKDNFKLMCTNAMY 60
    |||||||
Db 151 DSAFFSPVTDPIAGYSMIKHHPDFTMKKIKNNNDYOSIEELKDNFKLMCTNAMY 210
OY 61 NKPEITYYKAARKLLHSGMKILISOERIOSLKOSIDPMADLOKTRKOKDGTDSOGSDGG 120
    |||||||
Db 211 NKPEITYYKAARKLLHSGMKILISOERIOSLKOSIDPMADLOKTRKOKDGTDSOGSDGG 270
OY 121 CMOEREDSGDAEAHAFKSPSKENKKDKMDLEDFKSNLNR 163
    |||||||
Db 271 CMOEREDSGDAEAHAFKSPSKENKKDKMDLEDFKSNLNR 313

RESULT 4
O9BV48 PRELIMINARY: PRT; 459 AA.
AC 09BV48;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE SIMILAR TO BROMODOMAIN-CONTAINING 7 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISUE-BRAIN; NEUROBLASTOMA;
RC Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC001611; AAH01611.1; -.
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; bromodomain; 1.
FT NON-TER
SQ SEQUENCE 459 AA; 51714 MW; 9EF21D1454BCE63B CRC64;

Query Match 71.3%; Score 610; DB 4; Length 459;
Best Local Similarity 98.3%; Pred. No. 4.8e-43;
Matches 116; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 46 LKDNFKLMCTNAMYNKPEITYYKAARKLLHSGMKILISOERIOSLKOSIDPMADLOKTRK 105
    |||||||
Db 3 IQDNFKLMCTNAMYNKPEITYYKAARKLLHSGMKILISOERIOSLKOSIDPMADLOKTRK 62
OY 106 OKDGTDSOGSDGGCWMOREREDSGDAEAHAFKSPSKENKKDKMDLEDFKSNLNR 163
    |||||||
Db 63 OKDGTDSOGSDGGCWMOREREDSGDAEAHAFKSPSKENKKDKMDLEDFKSNLNR 120

RESULT 5
O9H8M2 PRELIMINARY: PRT; 501 AA.
AC 09H8M2;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE CDNA FLJ13441 f15, CLONE PLACE1002775, WEAKLY SIMILAR TO

```

DE PEREGRIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NX NCBI_TaxID=9606.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA Isegaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
 RA Tanase T., Nomura Y., Togliya S., Komai F., Hara R., Takeuchi K.,
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.;
 RT "NEO human cDNA sequencing project."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK023503; BABI4591.1; -
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; Bromodomain.1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00297; BROMO.1.
 DR PROSITE: PSS0014; BROMODOMAIN_2; 1.
 SQ SEQUENCE 501 AA; 55664 MW; 5C68F53097BA073C CRC64;

Query Match 43.98; Score 375; DB 4; Length 501;
 Best Local Similarity 47.58; Pred. No. 2e-23;
 Matches 77; Conservative 28; Mismatches 37; Indels 20; Gaps 3;

QY 1 DPSAFSPVDTLPAGYSMTIKHPDFTMKERIKNNQSYIELDKNFKTMNAMY 60
 DB 40 DPHGFAPVDTALAPGSMITIKHPMDGTMKDIVANEKTSVEFRADFLKCDNAMY 99
 QY 61 NKPEITYYKAKKLLHSGMKILSQERISLQSIDFNADIKTRKQKDGDTSGSGEDG 120
 DB 100 NRPTVYVYKLAKKILHAGFKMKSERLLATKRSMSPQDM-----DFSQQAALLG 149
 QY 121 CWQREKREDSGAEAHAHAFSP-----SKRNKKKDKMLDKRK 157
 DB 150 -----NEDTAVEEPYEPVPOVETAKSKRPSREVISCMEF 186
 RESULT 6
 Q9H7R9 PRELIMINARY; PRT; 233 AA.
 AC Q9H7R9;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE CDNA FLJ14330 FIS, CLONE PLACE4000261, WEAKLY SIMILAR TO
 DE PEREGRIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA Isegaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
 RA Tanase T., Nomura Y., Togliya S., Komai F., Hara R., Takeuchi K.,
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.;
 RT "NEO human cDNA sequencing project."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK024392; BABI4907.1; -
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; Bromodomain.1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00297; BROMO.1.
 DR PROSITE: PSS0014; BROMODOMAIN_2; 1.
 SQ SEQUENCE 233 AA; 26233 MW; FB96245BD86F4EB CRC64;

Query Match 38.48; Score 328; DB 4; Length 233;
 Best Local Similarity 67.18; Pred. No. 7.1e-20;
 Matches 57; Conservative 15; Mismatches 13; Indels 0; Gaps 0;

QY 1 DPSAFSPVDTLPAGYSMTIKHPDFTMKERIKNNQSYIELDKNFKTMNAMY 60
 DB 40 DPHGFAPVDTALAPGSMITIKHPMDGTMKDIVANEKTSVEFRADFLKCDNAMY 99
 QY 61 NKPEITYYKAKKLLHSGMKILSQE 85
 DB 100 NRPTVYVYKLAKKILHAGFKMKSQ 124
 RESULT 7
 Q9V1X2 PRELIMINARY; PRT; 861 AA.
 AC Q9V1X2;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE CG7154 PROTEIN.
 GN CG7154.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydriodea; Drosophilidae; Drosophila.
 NX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt R.G., Nelson C.R., Miklos G.L.G.,
 RA Artl J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,
 RA Borokova D., Botchan M.R., Bouck B.P., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davoport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Jengam C.,
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Paclob J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.C., Scheeler F., Smith H.,
 RA Shue B.C., Siding-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 RA Svrskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003618; AAE52557.1; -
 DR FlyBase: FBgn0031947; CG7154.
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; Bromodomain.1.

Db 191 DTYGYSPVPPEELPDYFEIKNPMDSITRNKLDGAYSTLEOFEDVFLICTNAMEY 250
 OY 61 NKPEIYYKAKKILHSGMKILISOERIQSLKOSID 95
 Db 251 NSADTVYRQAR-----AIQELAKDFENLRQDSD 280

RESULT 14

O9FIA2 PRELIMINARY; PRT; 145 AA.

ID O9FIA2
 AC O9FIA2:
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE DBJ|BA85417.1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 OX [1]
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX MEDLINE=99397451; PubMed=10470850;
 RA Kaneo T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
 RA Miyajima N., Tabata S.,
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
 RT Sequence features of the regions of 1,011,550 bp covered by seventeen
 RT P1 and TAC clones."
 RL DNA Res. 6:183-195(1999).
 DR EMBL: AB017059; BAB10578.1; -
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; bromodomain.1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00297; BROMO; 1.
 DR PROSITE: PS50014; BROMODOMAIN_2; 1.
 SO SEQUENCE 145 AA; 16649 MW; BD3E5CECF910CC63 CRC64;

Query Match 19.4%; Score 166; DB 10; Length 145;
 Best Local Similarity 45.3%; Pred. No. 1.3e-06;
 Matches 29; Conservative 16; Mismatches 19; Indels 0; Gaps 0;
 OY 6 FSPFVDTFAPGYSMTIKRHPSTKKEIKKNDYOSIELKDNFKLMCTNMIYKPEP 65
 Db 67 YAPVPPEELPDYFEIKNPMDSITRNKLDGAYSTLEOFEDVFLICTNAMEY 126
 OY 66 IYK 69
 Db 127 VYK 130

RESULT 15

O9SFX1 PRELIMINARY; PRT; 556 AA.

ID O9SFX1
 AC O9SFX1:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOHETICAL 62.1 KDA PROTEIN.
 GN FLSM4.12.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 OX [1]
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
 RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,
 RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;

RT "Arabidopsis thaliana chromosome 1 BAC FLSM4 genomic sequence."
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL, AC012384; AAF16663.1; -
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; bromodomain.1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00297; BROMO; 1.
 DR PROSITE: PS50014; BROMODOMAIN_2; 1.
 KW Hypothetical protein.
 SO SEQUENCE 556 AA; 62060 MW; 03878B1E71C891C2 CRC64;

Query Match 19.1%; Score 163.5; DB 10; Length 556;
 Best Local Similarity 27.6%; Pred. No. 9.3e-06;
 Matches 45; Conservative 28; Mismatches 55; Indels 35; Gaps 4;
 OY 1 DPSAFSPVDTFAPGYSMTIKRHPSTKKEIKKNDYOSIELKDNFKLMCTNMIYKPEP 56
 Db 161 DTYGYSPVPPEELPDYFEIKNPMDSITRNKLDGAYSTLEOFEDVFLICTNAMEY 219
 OY 57 AMIYKPEIYYKAKKILHSGMKILISOERIQSLKOSIDFMADLQTKRKQKDGTDPTSQSG 116
 Db 220 AMEYNSADTVYRQARAMELELAK-----DGNLRQESDGEPPVSL 261
 OY 117 EDGCGWQREDESDGDAEHAHAKSPKRNKKDKDKMLKFKSN 159
 Db 262 QOPKVKRGR-----PPSGGLKQLEOSLIDRTTSD 292

Search completed: July 11, 2002, 15:56:17
 Job time: 360 sec


```

Command line parameters:
-MODEL=framet.n2p.model
-Q/cgcn2.1/USPTO.spool/US09687230/runat.11072002.144407.29326/app_query.fasta_1.2401
-DB=pt_71
-QMT=fastlan
-SUFFIX=tp
-GAPDP=12.000
-GAREX=4.000
-MIMMATCH=0.100
-LOOPCL=0.000
-LOOPEXT=0.000
-GGAPDP=4.500
-OGAREXT=0.050
-XGAREXT=10.000
-XGAREXT=0.500
-FGAPDP=5.000
-FGAREXT=7.000
-YGAPDP=10.000
-YGAREXT=0.500
-DELOC=6.000
-DELEX=7.000
-START=1
-MATRIX=blosum62
-TRANS=human40.cdt
-LIST=45
-DOCALLIGN=200
-THR_SCORE=pct
-THR_MIN=100
-THR_MAX=100
-THR_MIN=0
-ALIGN=15
-MODE=LOCAL
-OUTFMT=pfs
-NORM_ext
-HEAPSIZE=500
-MINLEN=0
-MAXLEN=2000000000
-USER=US09687230.CGcn1.1.262
-NCPU=6
-ICPU=3
-LONGIO=
-DEV_TIMEOUT=120
-WARN_TIMEOUT=30
-NO_XLPHY
-WAIT
-THREADS=1

```

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
score_list:						
PIR2:J18845	+	517.50	602.17	3.0e-26	616	hypothetical protein CO1H6.7 -
PIR2:J123534	+	255.50	292.23	4.8e-09	715	hypothetical protein DKFZP434B0
PIR2:J062969	+	236.00	275.79	2.3e-08	1214	zinc finger protein, BR140 -
PIR2:J096791	+	236.00	271.45	4.9e-08	556	hypothetical protein F15M4.12
PIR2:J539162	+	228.00	249.03	3.6e-07	2440	transcription coactivator CREB
PIR2:J539161	+	227.50	248.44	3.9e-07	2441	CREB-binding protein - mouse
PIR2:J447371	+	217.50	238.10	1.7e-06	2068	transcription initiation factor
PIR2:J454277	+	216.00	234.98	2.2e-06	1414	transcription adaptor protein
PIR2:J539680	+	210.00	231.59	5.2e-06	1586	HBRM protein - human
PIR2:J5C5056	+	207.50	228.39	7.6e-06	1633	polydromo 1 - chicken
PIR2:J534036	+	206.50	228.82	8.6e-06	1559	hypothetical protein B0041.7 -
PIR2:J545251	+	205.00	225.77	1.1e-05	1572	SNF2alpha protein - human
PIR2:J533378	+	204.00	223.20	1.1e-05	1734	hypothetical protein F13C5.2 -
PIR2:J415522	+	201.50	223.10	1.8e-05	1332	probable transcription factor
PIR2:J568182	+	198.00	223.93	2.9e-05	757	probable transcription factor S
PIR2:J445282	+	194.00	212.39	5.9e-05	1647	SNF2beta protein - human
PIR2:J281845	+	193.00	218.32	6.2e-05	733	RING3 kinase - chicken
PIR2:J740964	+	191.50	218.64	7.3e-05	578	transcription factor btf1 homolog
PIR2:J153078	+	189.00	210.69	0.0001	1022	homeotic gene regulator - mouse
PIR2:J523733	+	186.50	204.43	0.0002	1490	DNA-binding protein TAF-II 256
PIR2:J567605	+	186.00	211.29	0.0002	638	hypothetical protein YD1070W -
PIR2:J144815	+	186.00	201.87	0.0002	1893	transcription initiation factor
PIR2:J566139	+	181.50	196.43	0.0004	754	female sterile homeotic (Fsh) h
PIR2:J566139	+	180.50	203.34	0.0004	522	hypothetical protein F2SP12.91
PIR2:J566139	+	176.50	201.85	0.0007	1613	hypothetical protein BRG1 - human
PIR2:J539059	+	173.50	188.41	0.0013	1613	protein BRG1 - human
PIR2:J218611	+	172.50	188.90	0.0015	1877	hypothetical protein F36F2.3 -
PIR2:J214137	+	172.00	199.10	0.0014	390	hypothetical protein C33G8.2 -
PIR2:J568181	+	170.00	191.25	0.0020	729	trifolidein - human
PIR2:J568181	+	169.00	189.26	0.0023	800	SART-1 protein - human
PIR2:J087378	+	169.00	189.14	0.0024	811	hypothetical protein DKFZP586E06
PIR2:J555959	+	168.50	190.02	0.0025	686	bromodomain protein BDP1 - yeast
PIR2:J213888	+	167.50	179.35	0.0035	3190	CREB-binding protein homolog -
PIR2:J562124	+	166.50	184.44	0.0035	990	protein T24D18.4 [imported] -
PIR2:J521424	+	166.00	179.54	0.0040	1618	nestin - human
PIR2:J218266	+	166.00	177.09	0.0042	2139	myosin heavy chain - Entamoeba
PIR2:J218266	+	165.00	187.75	0.0039	665	hypothetical protein PF80680V -
PIR2:J278937	+	165.00	180.57	0.0045	1287	hypothetical protein C52B9.8 -
PIR2:J228845	+	164.50	180.03	0.0049	1250	hypothetical protein F57C7.1a -
PIR2:J277401	+	164.50	178.56	0.0050	1479	transcription regulator WBSCR

seq_name: pir2:T18845

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C:Accession: T18845

R;Berks, M.

submitted to the EMBL Data Library, April 1996
A/Reference number: Z19030

A/Accession: T18845
A/Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA
A;Residues: 1-636 <WIL>

A; Cross-references: EMBL:Z71258; P
A: Experimental source: clone C01H6

A:Gene: CFSP-C01H6 7

A;Map position: 1
A;Tat: 30/1 - 109/3. 265/3. 451/3. 480/3. 535/3

C; Superfamily: bromodomain homology

F;172-227/Domain: bromodomain homology <BRU>

alignment_scores:

Quality:	517.50	Length:	711
Ratio:	1.387	Gaps:	25
Percent Similarity:	52.461	Percent Identity:	25.598

alignment_block:

```
US-09-687-230-1 x T18845 ..
Align seg 1/1 to: T18845 from: 1 to: 636
```

6 GGGGCATCGGCCCCGCCCGCCGCGCCGCCCTTGGCTTCGGGGCG . 53

4 GlytSerArgArgSerMetValGlyLeuProProThrArgArgAlaIar 20

54 . ශ්‍රේණිකාලයන්ගෙන් සෑදුනු පොදු කාලයක් සඳහා වෙනම ආවේණික අනුපාතිකයක් ලෙස ගත හැකිය.

```

30  acf wci vva sn thr pro Ser thra a thr pro Val Val 32
    |||:::  ||  ::  |||:::

```

[illegible]

|||::|||

103 AAGCGGCGCCGCACGGCCTTGGGCCGTGGGGGGGGGGCAAGCGGGGC 19

33 Proargininserala 3/

153 CGGTCGACATTGGCGAAGACACACAAGACACAAGTCGGACAAACACCT 20
||||: ::|||: |||||: ::::: |||||: ::::: |||||: :::::

38 Argala.....AtalysargVallystysGluInProGluGluGlu.. 51

203 CTACGACGAT.....GTAGAGAGCCCTTGAAGC 23

52 ...GIuAsPTyrLysAsnAsnSerAspProGluLysSerGluGlu 67

235 TGGTCTCAAGTAGGAGGGAACGACTACCGAATCTCCACGGGAGC 28

```

67 spcIuserGluIuserGlyaspIumethrThrProserArgIystr 83

```

285 TCCGGGCACGACTCCAGCCTCTTCGAAGACAAAAACGATCATGACAACA 33

```

      |||      :::::
84 ProgIvGIvA]agIv..... 89

```

335 CAAGCAAGAAAGCGAAAAAGCAAGAAAGGAGAGAAAGCAGATTTCCAG 38

[illegible]

30lyndyls.....lyndyls 20
31lyndyls.....lyndyls 20
32lyndyls.....lyndyls 20
33lyndyls.....lyndyls 20
34lyndyls.....lyndyls 20
35lyndyls.....lyndyls 20
36lyndyls.....lyndyls 20
37lyndyls.....lyndyls 20
38lyndyls.....lyndyls 20
39lyndyls.....lyndyls 20
40lyndyls.....lyndyls 20
41lyndyls.....lyndyls 20
42lyndyls.....lyndyls 20
43lyndyls.....lyndyls 20

```

98 eutrhspTyrHisLeuLysLysLysLysIleValAlaArgLysAlaAla 114
   |||
   |||
435 CGAGATGAGACCGGGTGAGAAATGAGCAGAAAAAGATCTCCAGTGTCA 484
   |||
   |||
115 ArgAspIleGluLys...GluLysGluValGluProGluValGluGluGlu 130
485 CGGCCCTGTGAGATTAAGCTTGCCTCTGAGAGC...CCTCTCAACA 528
   |||
   |||
130 uValProLysGluProThrProProProArgLysAlaProSerPheS 147
529 GCCTTTAGCCAAACAGAGAAGTAGACAGACACCCCTTCAAGAGCT 578
   |||
   |||
147 eSer.....TyrLeuProIleGluLeuMet 155
579 TTGAATCACTGATGAGACAATTCAGAGAAAAGATCCAAGCTCTTCTT 628
   |||
   |||
156 GluAspIleIleLeuArgLysLeuValGluLysAspProGluGluTyrPh 172
629 TTGATTTCCGTGACTGATTTTATGCTCTGGCTACTCCATGATCATTA 678
   |||
   |||
172 eAlaIleProValThrProSerMetAlaProAspTyrArgAlaIleIle 189
679 AACCCCAATGATTTTATGACATGAGAAAGAAAGATCAAGACAATGAC 728
   |||
   |||
189 yThrProMetAlaSerLysGluThrIleArgGluAsnIleGluAspGlyLys 205
729 TATCACTCCATGAGAACTAAAGATTAACCTTCAACTAATGCTACTAA 778
   |||
   |||
206 TyrAlaSerLeuProAlaMetLysGluAspCysGluLeuIleValSerAs 222
779 TGCCATGATTTACAATAACCCAGACCATTTATTTAAAGTCGCAAGA 828
   |||
   |||
222 nAlaIleGluIleTyrAsnGlnProAsnThrValPheTyrLeuAlaIleLys 239
829 AGCTGTTGACTCAGGAATGAAAAATTTCTTACGAGAGAAAGATTCAGAC 878
   |||
   |||
239 rGluSerAsnLeuIleAlaTyrTyrPheGlyGluGluIleTyrLeuAlaGpHe 255
879 CTGAGACAGACCTAGACCTTC..... 899
   |||
   |||
256 LeuPheIleSerLeuProMetAlaAsnLysIleProPheGluIleValGlu 272
900 .....ATGGCTGACTGCAGAAAACCTGA.....AACCA 930
   |||
   |||
272 yIleArgProLeuAlaProValProLysGluArgThrMetAlaLysArgL 289
931 AAGATGGAACAGACACCTCAGACAGTGGGAGAGACGAGCTGCTGGCAG 980
   |||
   |||
289 yAlaIleValLysAspGlyMetThrSerGluAsp....CysLeuGln 303
981 .....AGAGAGAGAGGACTCTGGAGATGCCGAC 1012
304 ValAlaAspProLysValArgLysGluLeuSerAlaLysLeuProGluAl 320
1013 ACAGCCCTTCAAGAGTCCAGCAAAAGAAATGAAAAAGAGCAAAAGATA 1062
   |||
   |||
320 a.....AsnAsnProLysAsnLysLysG 328
1063 TCCCTGAGAGATAGTTTAAACCAATTAATTAGAGAGAGAGAGAGCAG 1112
   |||
   |||
328 eGluLysLeuGluPheLeuSer..... 335
1113 CTTGACCCGATCGTGAAGAAATCTGGAGAAAGCTGACAGCGCTTGT 1162
   |||
   |||
336 .....GluLysAspGlyThrValIleuAsnValValA 346
1163 GAACAGTCAGTCCGAA.....TTTGAAGAGAAAAACAGATGGA 1203
   |||
   |||
346 lAlaIleLysAspSerGluAspGlyLysLeuGluAsnAlaProAlaGluVal 363
1204 CAAGCAGCTTGGGACTTCTCCATCTGTGGATCCCATTTGTAGAGAGCCA 1253
   |||
   |||

```

```

363 alThrIleGlyAspIleValGluProLeuGluGlu.....GlyThrPro 377
1254 GCCTACTGCTGCTGAGACGTGGAGATGACAACTGAGACCTGCTGG 1303
   |||
   |||
378 Gly.....MetIleGluMetAlaAspHisArgLeuPheSer.. 389
1304 AGTGAACTCTTTCGAGGGGTTCAAAAGAGATTAAGAGCAAAAGTCACTC 1353
   |||
   |||
390 .....GluAlaPro 392
1354 CAGCTTATATTTGAATTAATGAGCCCTACAGCTTCTTATGACCGCATTA 1403
   |||
   |||
392 roValAlaAsnTyrLeuAsnTyrGlyProTyrSerSerPheAlaProMetTyr 408
1404 GACTCCACATTTGCAAAATATCAGCAAGAGATTTCTGATTTAATCTATTC 1453
   |||
   |||
409 AspSerThrThrPalaThrMetThrLysGluAspThrAspLeuPheLeuAl 425
1454 AACCTATGGGGAAGACTGTGATCTTCCAACTGATTTCCAGATCCATGAGT 1503
   |||
   |||
425 gThrTyrGlyAspLysSerAsnAlaSerAspValMetSerMetArgArg 442
1504 TTTTGGCCAGTCCCAAGATTAATCCGATTCATGACAGATTAATTAATCTG 1553
   |||
   |||
442 heValGlyAspCysProGluPheSerGluIleIleGly...SerLeuLeu 457
1554 GATGTTTAAACAAAGAGGAGGACATTCAGAGCCCTCAAGAGATGAGAT 1603
   |||
   |||
458 AspThrLeuThrAspGlyGluHisSerLysThrMetLysGluLeuGlu.. 473
1604 GTCATTTGCTGAAAGATGAAAGGCCATCTAGACACTGACACAGAAAG 1653
   |||
   |||
474 .....AsnAlaGlyLysG 478
1654 AAATGGAGCAGATTTACAGAGTAGACCCACAGCGGCTTGGACCTCAGT 1703
   |||
   |||
478 luvAllyLysGlu.....GluValAspAsnAspGluTyrLysAsnGluThr 492
1704 ACTCAAGACAGGCTCATAGCGCTGAAGAGAGTAAACAAATTTGGCGCTCC 1753
   |||
   |||
493 ValLeuSerLeuIleAspAspValSerSerIleSerAsnLeuGlyIleG 509
1754 AGTTGAAGTTTGTGACTCTGAGAACCTGAATATTC..... 1790
   |||
   |||
509 uThrGlyPheLeuAsnAspIleArgGluGluValIleuValProAlaValG 526
1790 ..... 1790
526 luserAsnIleGluAsnAlaIleProGluPheMetAsnGluValAsnHis 542
1791 .....CAGAAGAACTGTATGAGACCCACAGATTCCTCAGGAGACT 1831
   |||
   |||
543 MetAsnValGlnGlnIleuAsnHisSerGlyGlnLysValLysAspRe 559
1832 CCAGGAAGCCCAAGATGAAGCTTTGAGCAGCAGACCCCTGGGAACATGA 1881
   |||
   |||
559 uAlaHisIleGlnGlnHisArgLeuValGlnGlnProProMetIleMet 576
1882 TCTGCTCTTGGGCTCCATCAGA..... 1906
   |||
   |||
576 eSerValGlnGlnLysAlaGlyGlnIleGluGlnLysLeuAlaLysLeu 592
1907 GAAATGACATCTGCTGACCAAGTGAACCAATTAAT 1939
   |||
   |||
593 GlnGlnHisLeuAlaHisGlnMetThrThrHis 603

```

seq_name: p12534

seq_documentation_block:
 hypothetical protein DKFZp434B094.1 - human (fragment)
 C:Species: Homo sapiens (hmn)
 C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 02-Sep-2000
 C:Accession: T12534

R.Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, June 1999
 A:Reference number: 217524
 A:Accession: T12534
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-715 <WAM>
 A:Cross-references: EMBL:AL080149
 A:Experimental source: adult testis; clone DKFZp434B094
 C:Genetics:
 A:Note: DKFZp434B094.1
 C:Superfamily: bromodomain homology
 F:113-168/Domain: bromodomain homology <BRO>

alignment_scores:

Quality: 255.50 Length: 369
 Ratio: 1.412 Gaps: 12
 Percent Similarity: 49.051 Percent Identity: 24.932

alignm_block:

US-09-687-230-1 x T12534

Align seg 1/1 to: T12534 from: 1 to: 715

```

302 CCTCTCGAAGACAAAAACGATCATGACAAACAGACAGACGGA 351
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 ProteuleuArgLeuGlnSer.SerLeuGlnSerGlnArgSer 33
352 AAAAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 401
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
33 InGlnArgGlnAspGlnGlnGlnGlnGlnGlnGlnGlnGln 49
402 AAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 434
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
50 TyrTrpGlnArgLeuArgHisAspLeuGlnArgLeuLeuLeu 66
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
435 .....CGAGATCGACACCGGCTGGAGATGAGCAGAAAGATC 477
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
66 uLeuLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 83
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
478 AGTGTACAGCCCGCTGTGATTAAGCTGCTCGAGAGAGCTCT 527
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
83 InValAlaMetGlnLeuArgLeu..... 90
528 AGCTCTTTAGCCAAAGAAAGAGAGAGAGAGAGAGAGAGAG 577
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
91 ..... 96
578 TTTGATCACTGATGACACATTCGACAGAAAGATCCAGTCTTCT 627
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
96 uLeuArgSerValLeuAspGlnLeuGlnAspGlnAspProAlaArg 113
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
628 TTTGATTTCTGCTGACGATTTTATTCGCTCGGCTCCATGATCAT 677
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
113 heAlaGlnProValSerLeuArgGlnValProAspTyrLeuAspHis 129
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
678 AAACACCAATGATTTAGTACATGAAGAAAGATCAAGACAAATGA 727
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
130 LysHisPrometAspPheAlaThrMetArgGlnGlnGlnGlnGln 146
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
728 CTATAGTCATAGAAAGAACTAAAGATTAACCTCAACCTGATAC 777
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
146 yTyrLysAsnLeuHisGlnPheGlnGlnGlnGlnGlnGlnGlnGln 163
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
778 ATGCCATGATTTACATAAAGACAGACACATTTATTAAGCTGCA 827
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
163 SerCysMetLysTyrAsnAlaArgAspThrValPheTyrArgAlaVal 179
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
828 AAGCTGTGCTCACTCAAGAAATGAATTTCTTACCGAGAAAGATTC 877
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
180 ArgLeuArgAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 193

```

```

878 CCTGAGCAGACATAGACTTCATGGCTGACTTGCAGAAACCTCGAAGC 927
193 ..... 193
928 AGAAATGAGACAGACACCTCACAGAGTGGGAGACGAGAGCTGCTGG 977
193 ..... 193
978 CAGAGAGAGAGAGAGACTCTGGAGATGCCGAGCA.....CAGC 1018
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
194 ...ArgGlnValAspSerIleGlyLeuGlnGlnGlnGlnGlnGlnGln 209
1019 CTTCAGAGCTCCCGCAAGAAATATAAAG.....AAAGACA 1056
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
209 uProGlnArgProAlaAlaAlaProArgArgProPheSerTrpGlnAsp 226
1057 AAGATATGCTTGAAGATTAAGTTTAAAGCAATATTA.....GAG 1097
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
226 aAspArgLeuLeuAspProAlaAsnArgAlaHisLeuGlnGlnGln 242
1098 AGAGAGCAGAGCAGCTTGACCGCATC.....GTGAGCA 1132
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
243 GlnLeuArgGlnLeuLeuAspMetLeuAspLeuThrCysAlaMetLys 259
1133 ATCTGAGAGAAAGCTGACACAG...CGGCTTGTGACAGATGCGCAT 1179
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
259 rSerGlySerArgSerLysAlaLysLeuLeuLysLysGlnLeuAla 276
1180 TTGAAGAGAAAGAAACAGANTGAAACAGAGCTGGGACTTCTCATCT 1229
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
276 euleuArgAsnLys.....LeuSerGlnGlnHisSer 286
1230 GTGATGCCATTTGAGAGAGCCAGCTACTGCTGTGAGACTGGGAAT 1279
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
287 ...GlnProLeuProThrGlyProGly..... 294
1280 GACAACGTGAAGACTTGAGTGTGAGTGAATATTTGACGGGGTTCAAG 1329
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
295 .....LeuGlnGlyPheGlnG 300
1330 AGGAT 1334
|||||
300 LuAsp 301
seq_name: p1r2:JC2069
seq_documentation_block:
zinc-finger protein, BR140 - human
N:Alternate names: bromodomain protein
C:Species: Homo sapiens (man)
C>Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 12-Sep-1997
C:Accession: JC2069
R:Thompson, K.A.; Wang, B.; Argaves, W.S.; Glancottl, F.G.; Schranck, D.P.; Ruoslahti
Biochem. Biophys. Res. Commun. 198, 1143-1152, 1994
A:Title: BR140, a novel zinc-finger protein with homology to the TAF250 subunit of TF
A:Reference number: JC2069; MUID:94151726
A:Accession: JC2069
A:Molecule type: mRNA
A:Residues: 1-1214 <THO>
A:Cross-references: GB:M91585
C:Comment: This is a nuclear protein with broad tissue distribution, but is especially
C:Superfamily: unassigned bromodomain proteins; bromodomain homology
C:Keywords: DNA binding; phosphoprotein; transcription regulation; zinc finger
F:653-708/Domain: bromodomain homology <BRO>
F:23,28,41,44/Binding site: zinc (Cys, Cys, His, His) #status predicted
F:120,205,462/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status
F:276,279,293,296/Binding site: zinc (His, Cys, Cys, Cys) #status predicted
F:301,304,317,320/Binding site: zinc (His, Cys, Cys, Cys) #status predicted
F:330,333,350,353/Binding site: zinc (Cys, Cys, His, Cys) #status predicted
F:386,389,401,405/Binding site: zinc (Cys, Cys, His, Cys) #status predicted
F:410,413,444,447/Binding site: zinc (His, Cys, Cys, His) #status predicted

```

alignment_scores: Quality: 245.50 Length: 347
 Ratio: 1.240 Gaps: 12
 Percent Similarity: 57.061 Percent Identity: 27.089

alignment_block:

US-09-687-230-1 x JC2069 ..

Align seg 1/1 to: JC2069 from: 1 to: 1214

```

325 ATGACAAACACAGACAGAAACCGGAAAAAGAAAAAGAGAGACA.. 372
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
524 lIethAsnArgLeuThrIleGlnArgLysSerGlnPheMetIleArgLe 540
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
373 ....ACGAGATTCAGGGGAGAAAGGAGAGAAAAAGAGAGAGATTAA 418
   ||| :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
540 uHlSerTyrTrpThrLeuLysArgLysSerArgAsnGlyValProLeuL 557
   ||| :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
419 GAGAGAT.....AAAAGAGCGAGATCGAGACCGGGTG 452
   ||| :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
557 euArgArgLeuGlnThrHisLeuGlnSerGlnArgAsnCysAspGlnVal 573
   ||| :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
453 GAGATGAGCGACAGA.....AAGATCTCCAGTGC 481
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
574 GlyArgAspSerGlnAspLysAsnTrpAlaLeuLysGlnIleuLysSe 590
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
482 TCACGCCCTTGAGATTAAGTACTTGCTCTCGAAGACCTCTCACAAGCT 531
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
590 rTrpGlnArgLysArgHisAspLeuGlnArgAlaArgLeuLeuValGlu 607
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
532 CTTTACCCAAACAGAA..... 548
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
607 euIleArgLysArgGlnLysLysLeuLysArgGlnThrIleLysValGlnGln 623
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
549 .....GAAGTAGACAGACAGACCCCTTCAAGAGACTTTGCAATCACT 589
   ||| :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
624 IleAlaMetGlnMetGlnLeuThrProPheLeuIleLeuLeuArgLysTh 640
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
590 GATGAGACAAATTCGACAGAAAGATTCAGAGTCTTTCTTCAATTTCCGTG 639
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
640 rLeuGlnGlnIleuGlnGlnLysAspThrLysAsnIlePheSerGlnPro 657
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
640 TGACTGATTTATATGCTGCTGCTACTCATCATGATCATTTAAACCCCAATG 689
   ||| :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
657 aLProLeuSerGlnValProAspTyrLeuAspHisIleLysLysProMet 673
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
690 GATTTTATGATCATGACAGAAAGAAAGATTCAGAACATGACTATCAGTCAT 739
   ||| :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
674 AspPhePheThrMetLysGlnAsnLeuGlnAlaTyrArgTyrLeuAsnPh 690
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
740 AAGAGAACTAAAGCATTAATCTCAACTAATGTGTACTAATGCCATGATT 789
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
690 eAspAspPheGlnGlnLysAspPheAsnLeuIleValSerAsnCysLeuLys 707
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
790 ACAATTAACSCAGACACATTTATTTAAAGTCGAAAGAGAGCTGGTGCAC 839
   ||| :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
707 yTAsnAlaLysAspThrIlePheTyrArgAlaIleValArgLeuArgGlu 723
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
840 TCAGCAATGAATAATCTTACCGACAGAAAGATTCAGAGCTTCAGACAGAG 889
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
724 GlnGlyGlyAlaValAlaArgGlnIleAlaArgGlnIleAlaGlyMetGln 740
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
890 CATTAACCTTCATGCGTGCATTCGCAAAAGTCAAGAGAGAAAGATGCAA 939
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
740 yLleAspPheGlnIleThrGlyMetHisIleProHisSerLeuAlaGlyAsp 756
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
940 CAGACACCTTCACAGATGGGGAGAGAGAGCTGCTGCAGAGAGAGAGA 989
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
757 GluAlaIleThrHisHisThr.....GlnAspAlaIleAlaGlnGlnArgLe 771
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
990 GAGGACTCTGAGAGATCCGAGACACACAGCTTCAGAGACTCCCGACAAAGA 1039
   ||| :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
771 uValLeuLeuGlnLysAsnGlnLysHisLeuPro..... 781

```

```

1040 AAATTAATAAAGAAAGACAAAGATATGCTTGAAGATTAAGTTAAAGCATATA 1089
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
782 ..ValGlnGlnGlnIleuLysLeuLeuLeu..GlnArgLeuAspGlnVal 797
   ||| :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
1090 ATTTAGAGAGAGAG.....CAGAGAGAGCTTGACCGCATCGTGAGAG 1130
   ||| :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
797 snAlaSerLysGlnSerValGlyArgSerArgArgAlaLysMetIleLys 813
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
1131 GAATCTGAGAGAAAGCTGACACAGCGGCTTGTAACAGTCACTGCGAATT 1180
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
814 LysGlnMetThrAlaLeuArgArgLysLeuAlaHis...GlnArgGlnThr 829
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
1181 TGAAGAGAGAAAGACAGAT.....GGACAAAGAGAGCTTGAGACTCTGC 1224
   ||| :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
829 rGlyArgAspGlyProGlnArgHisGlyProSerSerArgGlySerLeu 846
   ||| :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
1225 ATCCTGTGATCCCATTTGTAGAGAGCCAGGCTACTGC 1262
   ||| :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
846 hrPro.....HisProAlaIleLys 852

```

seq_name: pir2:D96791

seq_documentation_block:

```

hypoetical protein F15M4.12 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
R:Accession: D96791
R:History: A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzita
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: D96791
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-556 <STO>
A:Cross-references: GB:AE005173; NID:96554481; PIDN:AAF16663.1; GSPDB:GN001441
C:Genetics:
A:Gene: F15M4.12
A:Map position: 1

```

alignment_scores: Quality: 236.00 Length: 447
 Ratio: 1.078 Gaps: 20
 Percent Similarity: 48.993 Percent Identity: 24.161

alignment_block:

US-09-687-230-1 x D96791 ..

Align seg 1/1 to: D96791 from: 1 to: 556

```

172 AGCAGAGAGACAGAAAGTGGGCAAAACCTCTAGAGAGATATGTAGAG 221
   ||| :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
51 SerSerArgAsnProAsnPheSerAsnArgSerAsnAlaGlyArgAsn 67
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
222 AAGCCCTTGAAGCTGCTCCATAAGTAGAGAGAGAGAAAGTCAACCGAAT 271
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
67 eTAsnSerGlnAspAspAspArgLysArgArgAspLysLysHisArgLeu 83
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
272 CTCACGGGCAAGCTGGGGCAAGAGACTCCAGCTCTTGAAGACAAAGAG 321
   ||| :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
84 LeuHis.GlyLeuAsnSerHisGlnGly..... 92
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
322 ATCATGACAAACACAGAGACAGAAAGCGGAAAAAGAGAAAGAGAGAG 371
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::

```



```

1375 eLysSerArgPheValAspSerGlyGluMetSerGlySerPheProTyr 1391
1401 .....TATGACTCCAC 1411
1392 ArgThrLysAlaLeuPheAlaPheGluGluIleAspGlyValAspValGly 1408
1412 ATTTGCAATATCAGCAGAGATGTTGATTATC..... 1448
1408 sPhePheGlyMetHisValGlnAspThrAlaLeuIleAlaProHisGlnI 1425
1449 .....TATCAACCTATGGGAAACCTGTCTCCACAGT 1484
1425 IeGlnIlycysValTyrIleSerTyrLeuAsp..... 1435
1485 GATTTCAGCATCAGATTGTTGGCCACCTGC..... 1517
1436 .....SerIleHisPheArgProArgCysLeuArgThrAlaValTyr 1450
1518 .....CAGATTATCCGATG 1533
1450 rHisGluIleLeuIleGlyTyrLeuGluTyrValLysLysLeuValTyr 1467
1534 TCATGGCAGATAGTTTACTGGATGTTTAAACAAGAGG..... 1574
1467 alThrAlaHisIleTrrPAlaCysProProSerGluGlyAspPrrTyrIle 1483
1575 .....CATTCAGGACCCCTACA 1591
1484 PheHisCysHisProProAspGlnLysIleProLysProLysArgLysIle 1500
1592 AGAGATGGAGATGTCATTCGCTGAAGTGAAGGCCATACATAGACACTTG 1641
1500 nGluTrrPrrTyrLysLysMetLeuAspLysAlaPheAlaGluArgIleLea 1517
1642 ACACAGAGAAAAGAAATGAGACAGATTACAGAGTAGAGCCACGAGCGGT 1691
1517 sHisPrrTrrLysAspIle..... 1522
1692 TTGGACTCAGTACTCAAGACAGGCTCATAGCCTGAAGCAGTAGACAA 1741
1523 PheLysGlnAlaAsnGlnLysAspArgLysSerAlaLysGluLeuProTyr 1539
1742 TTTGGCGGTTCCAGTGAAGTTTGGACTCGAAGAGCTGAATATTCC 1791
1539 rPhe.....GluGlyAspPheTrrP 1546
1792 AGAGAAACTTGATGAGACACACAGATTGCTCAGAGAACTCAGAGAACCC 1841
1546 roAsnValLeuGluGlySer.....IleLysGluLeuGluGlu 1559
1842 CAGATGAAGCT.....TTGAGCACACAGACCC 1870
1560 GluGluGlnArgLysLysGluGlySerThrAlaAlaSerGluThrProG 1576
1871 TGGGAACATGATGTGTCTTGGGCTCATCAGAGAAATGACATCTTGC 1920
1576 uGlySerIle.....GlyAspSer..... 1582
1921 TGACAACTGACCAATATCTTAAAGATTGACAGCAAGTAAGTCAAGG 1970
1583 .....LysAsnAlaLysLysLysAsnAsnLys 1591
1971 TGATATCGTAAGCAGCTATGAGT...TCGAAGACATGGGATTTTCCA 2017
1592 LysThrAsnLysAsnLysSerSerIleSerArgAlaAsnLysLysLysP 1608
2018 TTCCTTCCCGCTCATGGAAAAACAATT 2045
1608 oSerMetProAsnValSerAsnAspLeu 1617
seq_name: plir2:S39161
seq_documentation_block:

```

CREB-binding protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 20-Apr-2000
C:Accession: S39161
R:Christia, J.C.; Kwok, R.P.S.; Lamb, N.; Hagihara, M.; Montminy, M.R.; Goodman, R.H.
Nature 365, 855-859, 1993
A:Title: Phosphorylated CREB binds specifically to the nuclear protein CBP.
A:Reference number: S39161; MIM:94019866
A:Accession: S39161
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2441 <CHR>
A:Cross-references: GB:S66385; NID:9435854; PUDN:AB28651.1; PUD:9435855
C:Superfamily: unassigned bromodomain proteins; bromodomain homology
F:1112-1169/Domain: bromodomain homology <BRO>

```

alignment_scores:
  Quality: 227.50      Length: 894
  Ratio: 0.625        Gaps: 42
  Percent Similarity: 40.716  Percent Identity: 20.470

alignment_block:
US-09-687-230-1 x S39161 ..

Align seg 1/1 to: S39161 from: 1 to: 2441

11 ATCGCGCGCGCGCGCGCGCGCG.....CCCCCTCGCGCGCGCG 54
838 LeuAlaProGlnAlaSerGlnLeuProCysArgProValThrGlnSerPr 854
55 GGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG..... 97
854 OleuHis.....ProThrProProAlaSerThrAlaGlyMetP 869
98 .....ACCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 130
869 roSerLeuGlnHisProThrAlaProGlyMetThrProProGlnProAla 885
131 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 166
886 AlaProThrGlnProSerThrProValSerSerGlyGlnThrProThrPr 902
166 ..... 166
902 oThrProGlySerValProSerAlaAlaGlnThrGlnSerThrProThrV 919
167 ..CAAGAAGCACAAAGACACAAAGTC.....GGACAAACACCTCTA 205
919 alGlnAlaAlaAlaGlnAlaGlnValThrProGlnProGlnThrProVal 935
206 CGAGGAGTATGTAGAAAGCCCTGAAGCTGTCTCTCAAACTAGAGAGGA 255
936 .....GlnProProSerValAlaThrProGlnSerSerGlnI 948
256 AGCAAGT.....CACGAACTCTCCACGCG.....CAGC 284
948 nGlnProThrProValHisThrGlnProProGlyThrProLeuSerGlnA 965
285 TCGGGGACAGACTCCAG.....CCTCTCGAAGCAAAAACGATCA 325
965 laAlaIleSerIleAspAsnArgValProThrProSerThrValThrSer 981
326 TGACAAACACAAAGACAG..... 343
982 AlaGluThrSerSerGlnGlnProGlyProAspValPrometLeuGluMe 998
344 ..AAACGGAAAAAGAGAAAGAGAGAGAGACAGATTCCAGGGAGAGAA 392
998 tLysThrGluValGlnThrAspAlaGluProLysProThrGlySer 1014
393 AAGGGAGAAAACGAGAAAGACTTAAGAGAT..... 425

```

1015 LysGlyGluProArgSerGluMetGluLysLeuGlnLysSer 1031
426 ..AAAAAGACGAGATCGAGACCGGTGAGATGAGCAAAAGATC 474
::: |||||::: ||| ::|||::: |||
1031 rGlnValLysGlnLysGlnLysGlnLysGlnLysSerGlnPro.... 1046
475 TCAGAGTCAAGCCCTGTGAGATGACTGCTCTGAGAAACCTCTC 524
::: |||||::: |||||::: |||||
1047 MetGluValGlnLysLysProGlu 1055
525 ACAGCTTTTACCAACAAGAGAGATGAGACAGACA..... 563
::: |||||::: |||||::: |||||
1056 ValLysValGlnLysGlnLysGlnLysGlnLysSerSerAsnSphTrAl 1072
563 563
1072 aserGlnSerThrSerProSerGlnProArgLysLysIlePheLysPro 1089
564 ..CCCTTCAAGAGCTTGAATCACTGATGAGCAATTCGAGACAAA 611
|||::: |||||::: |||||
1089 LcGlnLeuArgGlnAlaLeuMetProThrLeuAlaLeuArgGln 1105
612 GATCCAGTGCCTTTC...TTTCATTTCCCTGAGCTGATTTTAT... 653
|||::: |||||::: |||||
1106 AspProGlnSerLeuProPheArgGlnProValAspProGlnLeuLeu 1122
654 ..GCTCTGCTACTCATGATCATTAACACCCCAATGATTTAGTACCA 702
||| ||| |||::: |||||::: |||||
1122 ylleProAspTyrPheAspIleValLysAsnProMetAspLeuSerThrI 1139
703 TGAAGAAGAAAGATCAAGACATGACTATGCTCATAGACAACTAAAG 752
::: |||||::: |||||::: |||||
1139 IelyAsrGlyLeuAspThrGlyGlnLysGlnProThrGlnTyrVal 1155
753 GATTAACCTTCAAACTAATGTGACTAATGCCATGATTTACATTAACACA 802
|||::: |||||::: |||||
1156 AspAspValArgLeuMetPheAsnAsnAlaTrpLeuTyrAsnArgLysTh 1172
803 GACCATTTATTTAAAGCTGCAAGAGCTGTGCACTGAGAAAGAAA 852
::: |||||::: |||||
1172 rSerArgValTyrLysPheCysSerLysLeu.....AlaGluV 1185
853 TTCTTAGCCAGGAAGAAATTCAGACCTGAGACAGATGACTGATCATG 902
::: |||||::: |||||
1185 alPheGlnLysGlnLysLeuAspProValMetGlnSerLeuGlyTyrCys 1200
903 GCTGACTTGCAAAAACTCGAAAGCAAGAAAGATGAGACAGACC.... 947
::: |||||::: |||||
1201 CysGly.....ArgLysTyrGlnPheSerProGlnThrLeu 1213
947 947
1213 sCysTyrLysGlnLeuCysThrIleProArgAspAlaIleTyrLys 1230
948TCACAGAGTGGGAGAGAGGCTCTGG.....CAG 980
::: |||||::: |||||
1230 eTyrGlnAsnArgTyrHisPheCysGlyLysCysPheThrGlnIleGln 1246
981 AGAGAGAGAGAGACTGTGAGAGTCCGACACACAGCTTCAAGAGTCC 1030
|||::: |||||
1247 GlyGlnAsnValThrLeuGlyAsp.....AspPr 1256
1031 CAGCAAAAGAAATATAAAGAAAGAAAGATATGCTTGAAGATTAAGTTTA 1080
|||::: |||||
1256 oSerGlnProGlnThrThrIleSerLysAspGlnPheGlnLysLys...L 1272
1081 AAAGCAATATTTAGAG..... 1097
|||::: |||||
1272 yAsnAspThrLeuAspProGlnProPheValAspCysLysGlnCysGly 1288
1098 AGAGAGCAGAGACAGCTT..... 1115
|||::: |||||
1289 ArgLysMetHisGlnIleCysValLeuHisTyrAspIleIleTrpProse 1305

1116GACCGATCGTGAAGAAATCTGGA..... 1139
|||::: |||||
1305 rGlyPheValCysAspAsnLysLeuLysLysThrGlyArgProArgLys 1322
1140GMAAGCTGACACGCG 1157
1322 LuAsnLysPheSerAlaLysArgLeuGlnThrThrArgLeuGlnAsnHis 1338
1158 CTGTGTAACGTCAGTGCAGATTTGAAGAGAAACACAGATGGAACAC 1207
|||::: |||||
1339 LeuGlnAspArgValAsnLysPheLeuArgArgGlnAsn..... 1351
1208 GACGTTGGACTTCTCCATCCTGTGATCCCATGTTAGAGAGACGCT 1257
|||::: |||||
1352HisPro.....GluAlaGly 1357
1258 ACTGCTGTGAGACTGGGAATGACACTGGAAGACTGATGCTGAGAGTG 1307
::: |||||
1357 LuValPheValArgValAlaLysSer..... 1366
1308 AATACTTTCAGGCGTTCAAGAGATTAAGAGAAAGATCACTCCAGT 1357
|||::: |||||
1367AspLysThrValGlnValLysProGln 1375
1358 GTTA.....TATTGAATATATGGCCCTACAGTCTTATGACCGC 1398
::: |||||
1375 yMetLysSerArgPheValAspSerGlyGlnMetSerGlnSerPheProT 1392
1399 AT.....TATGACTCC 1409
1392 yArgThrLysAlaLeuPheAlaPheGlnLysLeuAspLysVal 1408
1410 ACATTTGCAATATATCAGCAAGAGATGATCTGATTAATC..... 1448
|||::: |||||
1409 CysPhePheGlyMetHisValGlnAspThrAlaLeuIleAlaProHisGln 1425
1449TATTCACCTATGAGGAGACTGATCTTCACA 1482
|||::: |||||
1425 nIleGlnLysValTyrIleSerTyrLeuAsp..... 1436
1483 GTGATTTCCATCCATGATGATTTGGCCAGCTGC..... 1517
1437SerIleHisPhePheArgProArgCysLeuArgThrAlaVal 1450
1518CAAGATTATCCGTA 1531
1451 TyrHisGlnIleLeuIleGlyTyrLeuGlnTyrValLysLysLeuValTyr 1467
1532 TGTCAATGCGATAGTACTGATGATGTTTAAACAAAGAGGG..... 1574
|||::: |||||
1467 rValThrAlaHisIleTyrAlaCysProSerGlnLysAspAspTyrI 1484
1575CATTCAGACCTTA 1589
1484 IephHisCysHisProProAspGlnLysIleProLysProLysArgLeu 1500
1590 CAAGAGATGAGAGATGCTCAATTCCTGAAGATGAGAGCCATCTAGACACT 1639
|||::: |||||
1501 GlnGlnTyrPyrLysLysMetLeuAspLysAlaPheAlaGlnArgIleI 1517
1640 TGACACAGGAAAGAAATAGAGCAGATTTACAGAGTACAGCCACAGGGC 1689
|||::: |||||
1517 eAsnAspTyrLysAspIle..... 1523
1690 GTTGGACTCCAGTACTCAAGACAGCTCATAGCCCTGAAGACGATAACA 1739
::: |||||
1524 ..PheLysGlnAlaAsnGlnLysPargLeuThrSerAlaLysGlnLeuPro 1539
1740 AATTTGGCGTTCCAGTTGAAGTTTTCAGTCCTGGAAGAGCGGAATATTT 1789
||| ||| |||::: |||||
1540 TyrPhe.....GlnGlyAspPheTr 1546


```

1614 ..... 1614
1001 AGATGCCGAGACACGGCTTCAGAGTCCAGCAAGAAATATAAAGA 1050
1615 ....ProGluSerThrProPheLeuLysPro.....ValAsnLysLysG 1628
1051 AAGACAAGATATG.....CTTGAAGAT 1073
1628 InValLysAspTyrTyrThrValLLeLysArgProMetLAspLeuGluThr 1644
1074 AAGTTAAAGCAATATTTAAGAGAGACGAGACGCTTGACCGCAT 1123
1645 lLeGlyLysAsnLLeGluAlaHisArgTyrHisSerArgLalGluTyrLe 1661
1124 CGTGAAGAAATCTGGAGAAAGCTGACGAGCGGCTTGGAACAGTCAGT 1173
1661 uAla.....AspLLeGluLeuLLeAlaThrAsnC 1671
1174 GCGAA.....TTTGAAGAGA 1190
1671 ySGLuGlnTyrAsnGlySerAspThrArgTyrThrLysPheSerLysLys 1687
1191 AAACCAAGATGGAACACACGCTGGAGCTTCTCCATCCTGTGATCCCAT 1240
1688 lLeuGluTyrAlaGlnThrGln..... 1695
1241 TGTAGAGAGCCGAGCTACTGCTGAGACTGGGAATGACACTGGAA 1290
1696 .....LeuLLeGluPheSerGlnHisCysGlyG 1705
1291 GACTTCAGCTGAGAGTAATCTTGCAGGGGTCAAGAGATAAAGC 1340
1705 lInLeGlnAsnAsnLLeAlaLysThrGln.....GluThrGlnArg 1718
1341 AACAAAGTCACTCCAGCTGTATATTTGAATATGGCCCTAAGTTCTTA 1390
1719 GluAsn..... 1720
1391 TGCACCGCATATATGATCCACATTTGCAAAATATCAGAAAGATGATCTG 1440
1721 .AlaProGluPheAspGlnAlaThrGlyAsn.....AspArgTyrA 1734
1441 ATTATATGATATTCACATATGGGGAAGACCTGATCTTCCAAAGATATTC 1490
1734 snPhe.....AspArgLysSerArgLalaserSerProLysp... 1746
1491 AGCATCATGAGTTTGGCCACGTCGCAAGATATCCGTATGATGATGCC 1540
1747 .....AspTyr..... 1748
1541 AGATAGTTACTGATGTTTAAACAAAGAGAGGATTCAGACGCTTAC 1590
1749 .....lLeAspValAlGluGlnHisGlyLysHisAlaserSerSera 1762
1591 AAGAGATGAGATGTCATTCCTGAGATGAAGGCCATCTAGACACTT 1640
1762 snSerLLeHisArgSerMetGlyAlaGlnAlaGlySerSerHisThrAla 1778
1641 GACACAGAAAGAAATGAGACGATACAGAAATGAGACGACGAGG... 1688
1779 ProAla.....ValArgLysProAlaProProGluPyr 1789
1688 ..... 1688
1789 cGlyLysValLysArgGlyArgGlyArgProAlaGlyGlnArgAspProV 1806
1689 .....CGTTTGACATCCAGTACTCAAGACAGGCTCATACGCTGAAA 1730
1806 alGluGlnAspLeuGlnCysSerThrAspAspGln..... 1817
1731 GCAGTAAACAATTTTGGCGTCCAGTTGAAGTTTGGACTCTGAGAACG 1780

```

```

1818 .....AspAspAspGluG 1822
1781 TGAATATTCAGAGAAACTTGATGAGACCAACCATTTGCTCAGGAGAC 1830
1822 uGlnAspPheGln.....GluV 1828
1831 TCCAGGAAGCCGAGAAATGAAAGCTTTGAGCAGACAGACCCCTGGAAATG 1880
1828 alSerGlnAspGlnAsnAlaAlaSerLLeuAspGlnGlyGluArg 1844
1881 ATC 1883
1845 lLe 1845
seq_name: p1r2:A54277
seq_documentation_block:
transcription adaptor protein p300 - human
C/Species: Homo sapiens (man)
C/Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 10-Dec-1999
C/Accession: A54277; S60344
R/Eckner, R.; Ewen, M.E.; Newsome, D.; Gerdes, M.; Decaprio, J.A.; Lawrence, J.B.; Li
Genes Dev. 8, 869-884, 1994
A/Title: Molecular cloning and functional analysis of the adenovirus E1A-associated 3
A/Reference number: A54277; MUID:95011587
A/Accession: A54277
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-2414 <ECK>
A/Cross-references: GB:001877; NID:g495300; PIDN:AAA18639.1; PID:g495301
A/Note: In the authors' translation 941-Ser is shown after 961 and consequently, real
R:Lundblad, J.R.; Kwok, R.P.S.; Laurance, M.E.; Harter, M.L.; Goodman, R.H.
Nature 374, 85-88, 1995
A/Title: Adenoviral E1A-associated protein p300 as a functional homologue of the tran
A/Reference number: S60344; MUID:95174889
A/Accession: S60344
A/Status: preliminary
A/Molecule type: protein
A/Residues: 552-660 <LDN>
C/Genetics:
A/Gene: GDB:EP300
A/Cross-references: GDB:9862958; OMIM:502700
A/Map position: 22q13.2-22q13.2
C/Superfamily: unassigned bromodomain proteins; bromodomain homology
C/Keywords: phosphoprotein; transcription; zinc finger
F:1075-1132/Domain: bromodomain homology <RBO>
F:89,507,1136,1295,1497,1834,1977,2062,2320/Binding site: phosphate (Ser) (covalent)
F:1734/Binding site: phosphate (Ser) (covalent) (by protein kinase A) #status predict
alignment_scores:
Quality: 216.00 Length: 740
Ratio: 0.663 Gaps: 33
Percent Similarity: 44.054 Percent Identity: 21.216
alignment_block:
US-09-687-230-1 x A54277 ..
Align seg 1/1 to: A54277 from: 1 to: 2414
17 CGCGCGCGC.....GCGCGCGCGCGCGCTGCGTGGCGCGCGCGG 54
|||||
870 ProProGlyProGlnSerGlnAlaLeuHisProProPro..... 882
55 GGTCTCGCGCGCGCGCGCTGCGCGCGCGCTGCGTGGCGCGCGCGGAA 104
|||||
883 ....ArgGlnThrProThrProThrProThrThrGlnLeuProGlnValG 898
105 GCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGG 154
|||
898 lInProSerLeuProAlaAlaProSerAlaAspGlnProGlnGlnInPro 914
155 GTGCGACATGGCGCAAGAACACAGAACGACCAAGTC..... 190

```


1692TTGGAC.....	1697
1468	yalsmeltleuaspyslalaValSerGIuagylIalValHisAspTryLys	1484
1698TCCAGTACTCCAGACAGCGCTCATACGCGTGAAGACAGT	1735
: :	
1485	AspIlePheIyGLNalAthrGIuaspArgLeuThrSerAlaIySGluLe	1501
1736	AACAATTTTGGCGCTCCAGTGTGAAGTTTTCACCTCAAAAGCTGAA	1785
1501	uProTrypHe.....GIuIyasp	1508
1786	TATTCAGAGAACTGTATGAGACACACAGATTCCTCAGGGAATCCAG	1835
: : :	
1508	heTrpProAsnValLeuGIuIyAsp.....IleIySGluIuLeu	1521
1836	GAAGCCCAAGATGACGT	1853
1522	GIuGIuGIuGIuGIuArg	1527

C>Date: 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change 20-Sep-1999
 C:Accession: JCS056; S60678
 R:Nicolas, R.H.; Goodwin, G.H.
 Gene 175, 233-240, 1996
 A:Title: Molecular cloning of polybromo, a nuclear protein containing multiple domains 1
 A:Reference number: JCS056, MID:97074677
 A:Contents: embryo
 A:Accession: JCS056
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-1633 <NIC>
 A:Cross-references: EMBL:X90849; NID:g951230; PID:CAA6235.1; PID:g951231
 A:Note: submitted to the EMBL Data Library, August 1995
 C:Comment: This is a nuclear protein with five repeats of a domain thought to be involved
 C:Genetics:
 A:Gene: pbl
 C:Superfamily: unassigned bromodomain proteins; bromodomain homology
 F:66-152/Domain: bromodomain repeat <BR1>
 F:73-128/Domain: bromodomain homology <BR01>
 F:198-284/Domain: bromodomain repeat <BR2>
 F:205-260/Domain: bromodomain homology <BR02>
 F:399-485/Domain: bromodomain repeat <BR3>
 F:406-461/Domain: bromodomain homology <BR03>
 F:537-623/Domain: bromodomain repeat <BR4>
 F:544-599/Domain: bromodomain homology <BR04>
 F:675-761/Domain: bromodomain repeat <BR5>
 F:682-737/Domain: bromodomain homology <BR05>

alignment_scores:

Quality:	207.50	Length:	698
Ratio:	0.638	Gaps:	33
Percent Similarity:	46.562	Percent Identity:	21.060

alignment_block:
 US-09-687-230-1 x JCS056 ..

Align seg 1/1 to: JCS056 from: 1 to: 1633

```

144 ACCGGGCGCGTCGACATGGCAGACAGACAGACAGATCGCA 193
      ||||| ||||| ||||| ||||| ||||| |||||
499 SerSerLAtHrSerApThrGlySer.....SerLysArgLysSerly 513
      ||||| ||||| ||||| ||||| ||||| |||||
194 CAAACACCTCTACGAGAGTATGTAGACAGCCCTTGACGTGCTCTCA 243
      ||||| ||||| ||||| ||||| ||||| |||||
513 sLyAsmEtArGlySGInArGmEtLySLeuTyAsnAlaValLeu. 529
      ||||| ||||| ||||| ||||| ||||| |||||
244 AAGTAGAGGAGACGACGACGACGACGACGACGACGACGACGAC 293
      ||||| ||||| ||||| ||||| ||||| |||||
530 .....GluAlaArgGluSerGlyThrGlyArgArgLeucys 541
      ||||| ||||| ||||| ||||| ||||| |||||
294 GACTCGACGCTCTTGAA...GACAAAAACGATCAT...GACAAACACA 337
      ||||| ||||| ||||| ||||| ||||| |||||
542 AspLeuPheMetValLysProSerLysLysAspLysProAspLysTyr 558
      ||||| ||||| ||||| ||||| ||||| |||||
338 G.....GACAGAAAG.....CGGA 351
      ||||| ||||| ||||| ||||| ||||| |||||
558 sLeIleuGlupPrometAspLeuLysMetLleGluHisnLleArg 575
      ||||| ||||| ||||| ||||| ||||| |||||
352 AAAAGAGAAAGAGAGAGAGATTCAGGGGAGAAAG..... 395
      ||||| ||||| ||||| ||||| ||||| |||||
575 snArpLySTyrValGluGluAlaMetLleAspAspMetLysLeuMet 591
      ||||| ||||| ||||| ||||| ||||| |||||
396 CGGAGAAAACGAGAGAGTAAAGAGATAAAAAGACGAGATCGAGA 445
      ||||| ||||| ||||| ||||| ||||| |||||
592 PheArgAsnAlaArgHisLysArgnGluGluLysArgnValTyAsnAs 608
      ||||| ||||| ||||| ||||| ||||| |||||
446 C.....CGGTGGAGATGAGCGCAAAAAGATCTCCAGT 480
      ||||| ||||| ||||| ||||| ||||| |||||
608 rAlaHisMetLeuGluLysLleLeuLysGluLysArgLysGluLeu... 623
      ||||| ||||| ||||| ||||| ||||| |||||
481 GTACAGCCCTGTGAGATTAGAC.....TTGCCCTCTGAGAAAGCT 521
      ||||| ||||| ||||| ||||| ||||| |||||

```

```

624 ....GlyProLeuProGluAspAspValAlaSerProLysLeuLys 638
      ||||| ||||| ||||| ||||| ||||| |||||
522 CTCACA.....ACCTCTTAGCCAAACAGAGAGAGTACAGACAC 565
      ||||| ||||| ||||| ||||| ||||| |||||
639 LeuSerArgLysSerGlyLleSerProLysLysSerLysTyrMetThr 655
      ||||| ||||| ||||| ||||| ||||| |||||
566 CTTTCAGAAAGCTTTGAATCACTGATGACAAATTCAG..... 605
      ||||| ||||| ||||| ||||| ||||| |||||
655 oMetGluGluLysLeuAsnGluValTyArgAlaValLysAsnTyThr 672
      ||||| ||||| ||||| ||||| ||||| |||||
606 .....AGAAAGATCCAAAGCTCTTCTTCATTCCTCGTACT 644
      ||||| ||||| ||||| ||||| ||||| |||||
672 sPlyAsArgLysArgLeuSerAlaLlePheLeuArgLeuProSerArg 688
      ||||| ||||| ||||| ||||| ||||| |||||
645 GATTTATTTGCTCCTGCTACTCCATGATCATTAACACCAATGATTT 694
      ||||| ||||| ||||| ||||| ||||| |||||
689 SerLysLeu...ProAspLysTyrLleThrLleLysLysProValAsp 704
      ||||| ||||| ||||| ||||| ||||| |||||
695 TAGTACCATGAAAGAAAGATCAAGAACATGACTATACGCTCAGAG 744
      ||||| ||||| ||||| ||||| ||||| |||||
704 tGluLysLleArgSerHisMetAlaAsnLysTyGlnAspLleAsp 721
      ||||| ||||| ||||| ||||| ||||| |||||
745 AACTAAAGGATAACTCAACTATATGTACTATGCGCATGATTACAT 794
      ||||| ||||| ||||| ||||| ||||| |||||
721 erMetValGluAspPheValMetMetPheAsnAlaCysThrTyAsn 737
      ||||| ||||| ||||| ||||| ||||| |||||
795 AAACCAAGACCATTTATTAAGCTCA.....AAGAGCT 832
      ||||| ||||| ||||| ||||| ||||| |||||
738 GluProGluSerLysLleTyLysAspAlaLeuValLeuHisLysVal 754
      ||||| ||||| ||||| ||||| ||||| |||||
833 GTTCACATCA.....G 843
      ||||| ||||| ||||| ||||| ||||| |||||
754 uLeuGluThrArgArgGluLleGluGluAspGluAspSerHisValPro 771
      ||||| ||||| ||||| ||||| ||||| |||||
844 GAATGAATATCTTACCCAGAGAAATTCAGACGACGAGACAGAT 893
      ||||| ||||| ||||| ||||| ||||| |||||
771 snValThrLeuLeuLleGluLleuLleHisnLeuPheValSerVal 787
      ||||| ||||| ||||| ||||| ||||| |||||
894 .....GACTTCATGCTCA 907
      ||||| ||||| ||||| ||||| ||||| |||||
788 MetSerHisGluAspAspGluGlyArgCysTyrSerAspSerLysAla 804
      ||||| ||||| ||||| ||||| ||||| |||||
908 C..... 908
      ||||| ||||| ||||| ||||| ||||| |||||
804 uLleProAlaValAspProAsnPheProAsnLysProProLeuThrPhe 821
      ||||| ||||| ||||| ||||| ||||| |||||
908 ..... 908
      ||||| ||||| ||||| ||||| ||||| |||||
821 sPleLleLleArgLysAsnValGluAsnAsnArgTyArgArgLeuAspLeu 837
      ||||| ||||| ||||| ||||| ||||| |||||
909 .....TTGCAGAAAACCTGAAAAGCAGAAAGA 934
      ||||| ||||| ||||| ||||| ||||| |||||
838 PheGluGluAsnMetPheGluValLeuGluValArgLysArgMet...As 853
      ||||| ||||| ||||| ||||| ||||| |||||
935 TGGAAACAGACCTCAGACAGAGTGGGAGAGCGAGCTGCTGGCAGAGA 983
      ||||| ||||| ||||| ||||| ||||| |||||
853 nArgThrAsp...SerGluLleTyArgLysAspAlaValGluLeuGlnP 869
      ||||| ||||| ||||| ||||| ||||| |||||
984 .....GAGAGAGAGACTCTGAGAT..... 1004
      ||||| ||||| ||||| ||||| ||||| |||||
869 hePheLleLysLleArgAspLleLeuCysLysAsnGlyGluLleLeu 885
      ||||| ||||| ||||| ||||| ||||| |||||
1005 .....GCCAAGCACACCTCTCAAGATCCAGACAA 1036
      ||||| ||||| ||||| ||||| ||||| |||||
886 SerProAlaLeuSerTyThrThrLysHisLysHisAsnAspValGlu 902
      ||||| ||||| ||||| ||||| ||||| |||||
1037 AGAAATATAAAAAGAA...GACAAAGATATGCTGAAGATTTAA 1083
      ||||| ||||| ||||| ||||| ||||| |||||
902 sGluLysLysGluLysLeuProLysGluLleGluLysAspLysLeuYSA 919
      ||||| ||||| ||||| ||||| ||||| |||||
1084 GCAATTAATTTAGAGAGAGACGAGACGAGCTTGACCCGATCGTAAGAA 1133
      ||||| ||||| ||||| ||||| ||||| |||||
919 rGluGluGluGluLysArgGluAlaGluLysSerGluAspSerSerLys 935
      ||||| ||||| ||||| ||||| ||||| |||||

```

```

1134 TCTGGAGAAAGCTGACGAGCGGCTTGTGAACAGTCACTGGCAATTGA 1183
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
936 ALaGlyLeuSerSerLeuHisArgThrTyrSerGlnAspCysSerPheLeu 952
1184 AGAAGAAAAACAGATGAGAACAGACGCTGGGACTTCCTCCACTCTGTGG 1233
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
952 sAsn.....SerMetTyrHisValGlyA 960
1234 ATCCCATTTGTAGAGAGCCAGCTACTGCTGCTGTAGACTGGGAATGACA 1283
      ||  ::::  |||||
960 sPTyValTyrValGluPro.....Ala 967
1284 ACTGGAAGACTTCAGTCTGAGAGTAATCTTGCAGGGGTTCAAGAGA 1333
      :::::  |||||  :::::  :::::  :::::  :::::
968 GluAlaAsnLeuGlnProHisIleValCysIleGluArgLeuTrpLys 984
1334 TAAAGGAACAAGTCACCTCCAGCTGTATATTGTAATTATGGCCCTACA 1383
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
984 pSerAlaGlyGlu.....LysTrpLeu..TyrGlyCysTrp. 995
1384 GTTCTTATGACCGCATATGACTCCACATTTGCAATATCAGACAGAT 1433
      |||  |||||  :::::  :::::  :::::  :::::
996 ..PheTyrArgProAsnGlnThrPheHisLeuAlaThrArgLysPheLeu 1011
1434 GATTCTGATTATATCTTCAACCTATGGGGAAGACTGTGATCTTCCAAG 1483
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1012 GluLysGluValPheLysSerAspTyrTyrAsnLysValProValSerLys 1028
1484 TGATTTCACG.....ATCCATGAGTTTGGCCACGT 1515
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1028 sIleLeuGlyLysCysValValMetPheValLysGluTyrPheLysLeu 1045
1516 GC.....CAAGATTATCCGTATGCTATGTCAGATAGTTTACTG 1553
      ||  :::::  |||||  :::::  :::::  :::::  :::::
1045 yProGlnAsnPheArgAspGluAspValTyrValCysGluSerArgTyr 1061
1554 GATGTTTAACAAAAGAGGCAATTCAGACCTTCAAGAAAGATGAGAT 1603
      :::::  |||||  :::::  :::::  :::::  :::::  :::::  :::::
1062 SerAlaLysThrLysSerPheLysLysIleLysLeuTrpThrMetProVa 1078
1604 GTCA.....TTGCCT..... 1613
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1078 lSerSerValArgPheValProArgAspValProLeuProValValArgV 1095
1614 .....GAAGATGAAGGCCATACT 1631
      :::::  |||||  :::::  :::::  :::::  :::::
1095 aAlaSerValPheAlaAsnThrAspLysAlaGluGluLysHisSer 1111
1632 AGGACACTTGACAGAGAAAGAAATGGAGCAGATTACGAAGTAGAGCC 1681
      |||||  :::::  |||||  :::::  :::::  :::::  :::::
1112 AspThrLeuAspAspSerLysValGlyGluSerIleLeu..... 1124
1682 ACCAGGCGCTTGAAGCTCAGTACTCAAGAGAGGCTCATAGCGCTGAAG 1731
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1125 .....HisLeuGluLysAspLysGluAsp..... 1132
1732 CAGTAACAATTTTGGCGTTCCAGTTGAAGTTTTCAGTCTGAA 1775
      |||||  |||||  :::::  :::::  :::::  :::::
1133 .....ValProValGluMetSerAsnGlyGlu 1141
seq_name: pLr2:T34036
seq_documentation_block:
  hypothetical protein B0041.7 - Caenorhabditis elegans
  C:Species: Caenorhabditis elegans
  C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
  C:Accession: T34036
  R:Fulton, R.; Wohlmann, P.
  Submitted to the EMBL Data Library, April 1997
  A:Description: The sequence of C. elegans cosmid B0041.
  A:Reference number: Z21466
  A:Accession: T34036

```

```

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1359 <FULL>
A:Cross-references: EMBL:AF000196; PIDN:AAC24256.1; GSPDB:GN00019; CESP:B0041.7
A:Experimental source: strain Bristol N2; clone B0041
C:Genetics:
A:Gene: CESP:B0041.7
A:Map position: 1
A:Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3

alignment_scores:
  Quality: 206.50      Length: 660
  Ratio: 0.599        Gaps: 23
  Percent Similarity: 52.273      Percent Identity: 21.212

alignment_block:
US-09-687-230-1 x T34036 ..

Align seg 1/1 to: T34036 from: 1 to: 1359

129 GCGCGGGGGGGCGGCGACCGGGCCCGGTGCGACATGGCGAAGACACA 178
      |||||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
86 ALaLysSerGluSerGluSerAspGluValAspGluArgLys 102
179 GAAGCAAGATCGGACAAACACCTTACGAGAGATGTAGAAAGCCCT 228
      ||||  |||||  |||  ::::  ::::  |||||
102 sLysSerLysSerLysLysValAspGlnLysLysGluLysSerL 119
229 TGAGCTGCTCTCAAGTAGAGGAAGCAAGTACCGCAACTGCCANG 278
      |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::
119 yLys.....LysArgThrTrp 124
279 GCGAGCTCGGGCGGCGACTCCAGCTCTTGAAGACAAAGATCATGCA 328
      ::::  |||||  ::::  ::::  ::::  ::::  ::::
125 SerSerSer.....GluAspGluAspSerAsp 134
329 CAACACAAAGACAGAAACGGGAAAAAGAAAGAAAGAGAAAGCAGA 378
      ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
134 uGluArgGluLysSerLysLysSerLysThrLysGlnTr 151
379 TTCACGGGGA.....GAAAAGGGGAAAAACGGGAAGATTAAAGG 422
      ::::  ||||  ::::  ::::  ::::  ::::  ::::  ::::
151 hrSerSerLysSerGluLysSerGluLysArgLysValLysLys 167
423 GATAAAAAGAACGAGATGAGAC.....CGGCTGAGAAATGAGCGC 463
      ::::  |||||  ::::  ::::  ::::  ::::  ::::
168 SerLysLysAsnLysGluLysSerValLysArgAlaGluThrSer 184
464 AGAAAAAGATCTCCAGTGCACGCCCTGTGAGATTAGACTTGCCCTG 513
      |||||  ::::  ||||  ::::  ::::  ::::  ::::
184 uGluSerAsp.....GluAspGluLysPro 193
514 AGAAGCTCTCAAGCTCTTACCAAAACAGAAAGAAATAGAACAGACA 563
      ::::  ||||  ::::  ::::  ::::  ::::  ::::
193 erLysLysSerLysGlyLeuLysLysLysAlaLysSerLysGlu 209
564 CCCCTTCAAGAGCTTGAATCAACTAGTACAGACAATTGCAGAAAAAGA 613
      ::::  ||||  ::::  ::::  ::::  ::::  ::::
210 SerGluSerGluAspGluLysGlu...ValLysLysSerLysLysSe 225
614 TCCAACTGCTTCTTTCATTTCCCTGTGACTGATTTTATGCTCTGCT 663
      ::::  ::::  ::::  ::::  ::::  ::::  ::::
225 rLysLysValValLysLysGluSerGluSerGluAspGluAlaProGlu 242
664 ACTCCATGATCATTTAAACCCCAATGATTTAGTACCTGAAGAAAG 713
      ::::  ||||  ::::  ::::  ::::  ::::  ::::
242 yLysLysThrGluLysArgLysArgSerLysThrSerSerLysGlu 258
714 ATCAAGAACAAATGACTATCATGCTACATAGAACATTAAGATTAAGTCA 763
      ::::  ||||  ::::  ::::  ::::  ::::  ::::
259 SerGluSerGluLysSerAspGluGluGluGluLysGluLysSerPr 275

```

```

764 ACTAATGTGTACTAATGCCATGATTACATAAACAAGACCATTTATT 813
      |||||
275 O.....LysProLysLysLysLysP 282
      |||||
814 ATAACTGCAAGAAAGCTGTG.....CACTACAGA 845
      |||||
282 roLeuAlaValLysLysLysSerSerAspLysLysSerLysLysSer 298
      |||||
846 ATGAAATTTCTTACGAGGAAATTCAGAGCTGAGCAGACAGATAGA 895
      |||||
299 ValGluValLeuProGluLysLysLysArgGlyAlaValThrLeuIle 315
      |||||
896 CTTCATGGCTGACTGCAGAAACTCGAAAGCAAGAAATGAAACAGACA 945
      |||||
315 rAspSerGluAspGluLysAspGluLysSerLysSerLysLysSer 332
      |||||
946 CCTCACAGATGGGAGAGACGAGCGCTGCGCAGAGAGAGAGAGAGC 995
      |||||
332 ALGluGluLysValSerLysLysLysAlaLysLysGluLysSerGlu 348
      |||||
996 TCTGAGATGCCGAGACACAGCGCTTCAGAGATCCGACAAAGAAATAA 1045
      |||||
349 SerGlySerAspSerSerSerGluLysSerLysLysValAsnArgLys 365
      |||||
1046 AAGAAAGACAAAGATATGCTTGAAGATAAG.....TTTAAAGCA 1086
      |||||
365 sLysLysGluLysProGluLysLysLysGlyLysLysLysLysLys 382
      |||||
1087 ATATATTAGAGAGAGACGAGAGCGCTGACCGCATCGTGAAGATCT 1136
      |||||
382 eLysLysGluLysGluThrLysAspAlaGluAlaGluLysGluArg 398
      |||||
1137 GAGAGAAAGCTGACACGCGCTTGTGACAGTCAGTCGCAATTTGAAAG 1186
      |||||
399 .....ArgLysArgLysGluLysLysGluLysLysLysLysLys 409
      |||||
1187 AAGAAACCAAGATGAAACACGAGCTTGGACTTCTCAGCTCTGTGATC 1236
      |||||
410 .....AsnG 411
      |||||
1237 CCATTGTAGAGAGACGAGCTTACTGCTGTGAGAGTGGATGACAAC 1286
      |||||
411 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 425
      |||||
1287 GGA.....AGACTTCAGTCTGAGTGAATTAATCTTGGCAGGG 1321
      |||||
426 GlyThrSerSerGluArgLysLysSerValValLeuAspProAsp 442
      |||||
1322 GTTCAAGAGATAAAGAGAAAGCAAGTCACCTCCGTTATATTGAT. 1370
      |||||
442 rSerThrValAspGluLysLysLysLysLysLysLysLysLysLys 458
      |||||
1371 .....TATGGCCCTACAGTCTTATGACCCGAT..... 1400
      |||||
458 eLysValArgLysLysLysProLysGluAlaHisGlyLysLysLys 474
      |||||
1401 TATGATCCACATTTGCAATATACAGAGATGATTTGATTTAATGTA 1450
      |||||
475 TyrAspCysAlaPheGluSerLysAspArgLysSerThrGluLys 491
      |||||
1451 TTCAACSTATGGGAGAAAGCTGATGCTCAAGATGATTTGACATGAT 1500
      |||||
491 yGluLysLeuAlaHisCysMetGlyLysLysLysLysLysLysLys 508
      |||||
1501 AGTTTGTGSCASTGSCAAGATTTATCCGTATGTC.....ATGSCAGAT 1544
      |||||
508 hrPheLeuHisThrValLeuMetHisGluLysLysLysLysLysLys 524
      |||||
1545 AGTTTGTGATGTTTAAACAAAGAGAGGATTCAGACCCCTACAGA 1594
      |||||
525 ArgValLeuValValValProLysAsnValLysLysLysLysLys 541
      |||||
1595 GATGAGATGTGATTTGCTGAGAGATGAGGCCATGACAGACATTTGACA 1644

```

```

      |||||
541 uphGluLysThrPheValAspAsnAspGluGluLeuAspThrLysAsp 558
      |||||
1645 CAGGAAAGAAATGAGACAGATTAACAGAAATAGAGCCAGCGCGCTTG 1694
      |||||
558 AlaAsnGlu.....Leu 561
      |||||
1695 GACTCC.....AGTACTCAGACAGCGCTCATGCGCTGAAACA.... 1733
      |||||
562 AspSerLysThrLysThrLysLysAspArgArgAlaLeuLysAlaTrp 1764
      |||||
1734 .....GTAAACAAATTTGGCGTTCCAGTGAAGTTT 1764
      |||||
578 sSerSerLysThrProSerValMetLysLysLysLysLysLysLys 1937
      |||||
595 LeuThrThrValGluAspAspProLysLysLysLysLysLysLys 611
      |||||
1815 AGATTGCTCAGGGAAGCTCCAGGAAAGCCAGAAATGAAAGCTTTGAG 1864
      |||||
612 ArgArgLysGluLysAlaLysGluAspPheArgLysThrLysLys 1899
      |||||
1865 ACCGCCGCGGAGAACATGATCTGT.....CTCTGGGCTCCCT 1937
      |||||
628 OGlyProAspMetValValCysAspGluAlaHisLysLysLysAsn 645
      |||||
1900 CATCA.....GAGAAATGATC.....TTGCTGAACAGATGACCAATA 1937
      |||||
645 sPserAlaLeuSerLysCysMetValLysLysLysLysLysLysLys 661
      |||||
1938 ATCTTAAAGAAATTCACACAGCAAGTAATCTCC 1967
      |||||
662 CysLeuThrGlyThrProLeuGluAsnAsn 671
      |||||
seq_name: p1r2:S45251

```

seq_documentation_block:

```

SNF2alpha protein - human
C:Species: Homo sapiens (man)
C:Date: 10-Dec-1994 #sequence_revision 10-Nov-1995 #text_change 28-May-1999
C:Accession: S45251
R:Chiba, H.; Muramatsu, M.; Nomoto, A.; Kato, H.
Nucleic Acids Res. 22, 1815-1820, 1994
A:Title: Two human homologues of Saccharomyces cerevisiae SNF2/SNF2 and Drosophila br
A:Reference number: S45251; MUID:94268902
A:Status: preliminary
A:Accession: S45251
A:Molecule type: mRNA
A:Residues: 1-1572 <CH>
A:Cross-references: GB:ID26155; NID:q505086; PIDN:BA05142.1; PID:di1005684; PID:998766
C:Superfamily: unassigned bromodomain proteins; bromodomain homology
F:1409-1464/Domain: bromodomain homology <BRO>

```

alignment_scores:

```

Quality: 205.00 Length: 349
Ratio: 1.265 Gaps: 14
Percent Similarity: 46.418 Percent Identity: 23.496

```

alignment_block:

```

US-09-687-230-1 x S45251 ..

```

```

Align seg 1/1 to: S45251 from: 1 to: 1572

```

```

321 GATCATGACAAACACAGAGAC.....AGAAAGCGGAAAGAG 358
      ||| |||||
1257 AspMetAspArgArgArgGluAspAlaArgAsnProLysArgLysPro 1273
      ||| |||||
359 AAGAAAGAGAGAGAGAGATTCAGAGG..... 386
      ||| |||||
1273 GLeuMetGluGluAspGluLeuProSerThrLysLysAspAspAlaG 1290

```

```

387 .....GAGAAAG..... 395
1290 luValgluArgleuthrCysgluIngluIngluysIlephglYarg 1306
396 ...GGGAGAAAAGGAGA..... 410
1307 GlySerArgInlArgArgAspValAspTyrSerAspAlaLeuthrGlu 1323
410 ..... 410
1323 sglInlPleuArgAlaIlegluAspGluAsnLeuIngluMetgluInglu 1340
411 ....AGAGTTAGGAGGATAAAAAGAGCA.....GATCAGACCGG 449
1340 luValAlrLeuLysLysArgglYargArgArgAsnValAspLysAspPro 1356
450 GTGGAGATGAGCAGAAAAAGATCTCAGGTGCAGCCCGCTGTGAGAT 499
1357 AlAlYsgluAspValgluLys.....AlAlYsgLysArgArg 1368
500 AGACTTGCTCCTGAGAGAGCCTCTCACAGCTCTTACGCAACAGAGAG 549
1368 gglYargProAlaglulYsLeuSerProAsnProProlYs..... 1382
550 AAGTAGACAGACACCCCTTACAGAGCTTTGATCAACAGTGATG...AGA 596
1383 .....LeuthrLysgluMetAsnAlaIleIleAspThr 1393
597 CAATTGCAGAAAAAGATCCAGT.....GCTTCTT 628
1394 ValIleAsnTyrLysAspSerSerGlyArgGluLeuSerGluValAlpheI 1410
629 TTCATTCTCTGACGACTGATTTATGCTGCTGCTCTCCTCAGATCATTA 678
1410 egluLeuProSerArgLysgluLeu...ProgluTyrTyrGluLeuIleA 1426
679 AACACCCAGATGATTTACTACATGAAGAAAAAGATCAAGAACATGAC 728
1426 rglYserProValAspPheLysLysIleLysgluArgIleArgAsnHISLys 1442
729 TATCAGTCCATGAAAGACTAAAGATTAACCTCAACTAATGTGATTA 778
1443 TyrArgSerLeuGlyAspLeuInlYsAspValMetLeuLeuCysHISAs 1459
779 TGCATGATTACATTAACAGACGACCTTATATATAAGCTGCAAGA 828
1459 nAlaeglInlPheAsnLeuGlySerGlnIleTyrGluAspSerIleV 1476
829 AGCTGTGCACTCAGGAATGAATAATCTTACGCAAGAAAGATTCAGAC 878
1476 allEu..... 1477
879 CTGAAGCAGAGATGACTTCTGCTGCTGCTGCAAAAACTGCAAGCA 928
1478 .....GlnSerVal...PheLysSerAlaArgGlnLysIleAlaLysgl 1491
929 GAAAGATGAGACAGACACTCAGAGTGGGAGGAGGAGGAGGCTGCTGC 978
1491 ugluInlusergluAspGluSerAsnGluInglu.....G 1503
979 AGAGAGAGAGAGAGCTGTGAGATGCCAGAGCAGACGCTTCAAGAGT 1028
1503 luGlulgluAspGluInlusergluSerGluAlaLysSerValLysVal 1519
1029 CCCAGCAAGAAAAATAAAAAGAAAGCAAAAGTATGCTTGAAGATTA 1078
1520 LysIleLysLeuAsnLysLysAspAspLysglYargAspLysglYLysgl 1536
1079 TAAAGC.....AATATTTAG 1095
1536 yLysLysArgProAsnArgglYLysAlaLysProValAlaSerAspPheA 1553
1096 AGAGAGACAGAGAGAGCTTACCGCATCTGAAGAAATCTGAGAGA 1142

```

```

seq_name: p1r2.T33328
1553 spSerAspGluIngluAspGluArgGluIngluInlserGluLysSerGly 1568
seq_documentation_block:
  hypothetical protein F13C5.2 - Caenorhabditis elegans
  C:Species: Caenorhabditis elegans
  C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 02-Sep-2000
  R:Wohlmann, P.; Hawkins, J.; Gillingham, B.
  Submitted to the EMBL Data Library, July 1998
  A:Description: The sequence of C. elegans cosmid F13C5.
  A:Reference number: 221324
  A:Accession: T33328
  A:Status: preliminary; translated from GB/EMBL/DBJ
  A:Molecule type: DNA
  A:Residues: 1-374 <MO>
  A:Cross-references: EMBL:AF077531; PIDN:AAC64610.1; GSPDB:GN00028; CESP:F13C5.2
  A:Experimental source: strain Bristol N2; clone F13C5
  A:Genetics:
  A:Gene: CESP:F13C5.2
  A:Map position: X
  A:Introns: 25/3; 135/3; 189/3; 313/1
  C:Superfamily: bromodomain homology
  F:141-198/Domain: bromodomain homology <BRO>

alignment_scores:
  Quality: 204.00 Length: 365
  Ratio: 1.041 Gaps: 18
  Percent Similarity: 53.699 Percent Identity: 26.849

alignment_block:
  US-09-687-230-1 x T33328 ..

Align seg 1/1 to: T33328 from: 1 to: 374

256 ACGAAGTCACCGAAGCTCTCCACGGGAGCTCGGGCAGCACTCCAGCTC 305
||||| ||| ||| ||||| ||||| ||||| ||||| |||||
4 ThrArgSerAlaAsnHisProAlaProLysArgAlaIleSerProAspSe 20
306 TTCG.....AAGCAAAAAAGATCATGCAAAACA 334
||||| ||||| ||||| ||||| ||||| ||||| |||||
20 rAspAspGluIngluAspAspPheLysSerLys..... 31
335 CAAGACAGAAAGCGAAAAAGAGAAAGAGAGAGAGAGATTCAG 384
||||| ||||| ||||| ||||| ||||| ||||| |||||
32 ..ArgThrProSerGly..ArgGlyArgGlyArgGlyArgGlyArg 46
385 GGGAGAAAAAGGGA.....GAAAAAG 407
||||| ||||| ||||| ||||| ||||| ||||| |||||
47 GlyArgGlyArgGlyGlyAlaThrAlaAlaGlyAlaThrValGluGly 63
408 AGAAGATTAGAGGATTAAGAAAGAGAGAGAGAGAGAGAGAGAGAG 457
||||| ||| ||| ||||| ||||| ||||| ||||| |||||
63 yLysSerThrArg.GlyGlyGlyThrGlyArg.....GlyGlyArg 76
458 TGAGGCAAGAAAAAGATCTCAAGTGCAGCCCGCTGAGATTAAGCTGC 507
||||| ||||| ||||| ||||| ||||| ||||| |||||
77 ..GlyArgProArgGly.SerLysAsnSerAspGlyGlyAlaGlyGly 92
508 CTCTGAGAGAGCTCTCAGAAAGCTCTTACGCAACAGAA..... 548
||||| ||| ||| ||||| ||||| ||||| ||||| |||||
92 roProAlaLysArgGlyArgGlyLysLysAlaLysSerGluSerGluAsp 108
549 GAGGTAGAACAGACACCCCTCAGAAAGCTTGAATCACTGATGAGAGCA 598
||||| ||||| ||||| ||||| ||||| ||||| |||||
109 GluAlaGluSerAspHisLeuHisAspGluLysLysCysLeuSerI 125
599 ATTGCAGAAAAAGATCCAGTCTTC.....TTTCAATTCTCTGNG. 641
||||| ||||| ||||| ||||| ||||| ||||| |||||
125 eleuLysgluPhegluLysSerThrHisAspSerPheThrPheProPheA 142

```



```

308 yrhIs...serpHeGlutyrAspLysGluThrMetIleLysArgLeuLys 323
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
357 AAAAAAGAAAGAGAGACAGATTCACAGCGCAAGAAAGGAGCAAAAGC 406
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
324 LeuGluGluSerAspLysMetIleGluLysGluLysLysLysSer 340
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
407 GAGAAGAGTTAAGAGAGATAAAGAGAGAGAGATGAGACCGGGTGAGA 456
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
340 gSerAspLeuGluAlaIleThrAspLeuGluAsnAspArgGluAsnThrAsnA 357
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
457 ATGAGCAGAAAAAGATCTCCAGTGTACGCCCCCT..... 491
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
357 spGluProAspThrAsnGluLysLeuProThrProGluLysSerThrPhe 373
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
492 .....GTGAGATTAGCTTGCTCC 511
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
374 SerAspThrGlyAsnLysArgProLysGlnSerAsnLeuAspLeuThrVa 390
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
512 TGAG.....AGCCTCTCACAGCTCTTAG 537
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
390 LAsnLeuGluIleGluAsnLeuSerLeuLysHisLeuSerSerIleG 407
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
538 CCAAAACA..... 545
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
407 LngLlnLysSerGlnLeuGluIleSerAspTrgluLeuLysHisLeu 423
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
546 .....GAAGAACT 553
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
424 IleMetAspValArgLysAsnArgSerLysTrpThrSerAspGluArgIle 440
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
554 AGACAGACACCCCTTCAAGAACTTTGAATCACTGATGAGACAATTGC 603
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
440 eGluGlnGluLysLeuLysGluAlaCysGluLysValValLeuGluLeuA 457
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
604 AAGA...AAAGATCCAGTGTCTTCTTCACTTTCCTGTGACTGATT 650
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
457 rGAsnTrpThrGluHisSerThrProPheLeuAsnLysValSerLysArg 473
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
651 ATTGCTCTGCTACTCCATGATCATTAACACCACTGATTTAGTATAC 700
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
474 GluAlaProAsnTrpHisGlnIleIleLysLysSerMetAspLeuAsnH 490
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
701 CATGAAGAAGAAATCAGACAAATGACTATCATGCTCATAGAAGAACTAA 750
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
490 rValLeuLysLysLeuLysSerPheGlnTrpAspSerLysGlnGluPheV 507
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
751 AGGTAATCACTCAACTAATGTGTACTAATGCATGATTACATAATAACA 800
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
507 aLAspAspIleMetLeuIleTrpLysAsnCysLeuThrTrpAsnSerAsp 523
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
801 GAGACATTATTATAAAGCT...GCAAGAAGCTGTTCACACTCAAGAT 847
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
524 ProSerHisPheLeuArgGlyHisAlaIleAlaMetGlnLysLysSerIe 540
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
848 GAAATTTCTTACCCAGAAAGAAATTCAGAGCTGAAGCAGACATAGACT 897
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
540 uGlnLeuIleArgMetIleProAsnIleThrIleArgAsnArgAlaAspL 557
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
898 TCATGGCTCACTTGCAAAACTCGAAAGCAGAAAGAT..... 935
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
557 euGluLysGluIleGluAspMetGluLysAspLysAspTrpTrgluLeuAsp 573
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
936 ...GGAACAGACCTTCACAGAGTGGGAGAGACGA.....GGCTG 973
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
574 GluGluGluGluValAlaGlySerGlyArgLysGlyLeuAsnMetGlyAl 590
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
974 CTGGCAGAGAGAGAGAGAGACTCTGGAGATCCGAGACACAGCCTTCA 1023
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
590 aHisMetLeuAlaLysGluAsnGlyLysValSerGluLysAspSerSerL 607
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1024 AGAGTCCCGCAAGAA..... 1040
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

```

```

607 ySThrValLysAspGluAlaProThrAsnAspAspLysLeuThrSerVal 623
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1041 .....AATAAAAGAAAGCAAA..... 1058
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
624 IleProGluGlyGlnLysGlnLysAspLysThrAlaSerSerThrValTh 640
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1059 .....GATATGCTTGAAGATTAAGTTTAAAGCAATAATTAGAGA 1098
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
640 rValHisGluAsnValAsnLysAsnGlnIleLysGluAsnGlyLysAsnG 657
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1099 GAGACAGAGACAGCCTTGACCCCATCGTCAAGACATCTGCG.....AGG 1141
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
657 LngLlnGlnAspMetValGluGlnSerSerLysThrGluAspSerSerLys 673
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1142 AAAGCTGACGACGCGCTTGTAACAGTCAGTCGCAATTTGAAAGAA 1191
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
674 AspAlaAspAlaAlaLysLysAspThrGluAspGlyLeu...GlnAspLys 689
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1192 AACCAAGATGAACAAGACAGCTTGGAGCTTCTCCATCTGTGATCCAT 1240
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
689 sThrAlaGluAsnLysGluAlaGlyGluAsnAsnGlnGluGluGluAspA 706
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1241 .....TGTAAGAGACCCAGCGTA 1258
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
706 sPAspAspGluAspGluAspGluAspMetValAspSerGlnSerTrpLeu 722
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1259 CTGCTGTGAGAGCTGGAGATGACAACTGAGACCTGACGTGCG..... 1303
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
723 LeuGluLysAspAspAspArgAspAspLeuGluIleSerValTrpLysTh 739
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1304 .....AGTGAATACT 1313
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
739 rValThrAlaLysValArgAlaGluIleCysLeuLysArgTrpGluTrp 756
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1314 TT.....GCAGGGGTCAAGAGGA 1333
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
756 IleLysAsnGlyLysLeuAsnSerAspSerGluAlaPheLeuLysAsnTr 772
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1334 TAAAGCAACAAGTCACTCCAGTGTATATTGAAATTAGGCGCCCTCA 1383
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
772 oGlnArgMetLysArgPheAspGlnLeuPheLeuGluTrpLysGluGlnL 789
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1384 GTTCTTATGACACCGCATTTGACTCCACACTTTGCAAT..... 1421
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
789 yAlaLeuGlnSerTrpArgGlnLysIleGluGlnAsnSerIleMetLys 805
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1422 .....ATCAGCAAGATGATTCGATTATATC..... 1448
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
806 AsnGlyPheGlyThrValLeuLysGlnGluAspAspAspGlnLeuGlnP 822
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1449 .....TATTCACCTATGGGGAAGACTCT.....GATCTTCCA 1482
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
822 eHisAsnAspHisSerLeuAsnGlyAsnGlnAlaPheGluLysGlnPhe 839
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1483 GTGATTTCAGCATCCATGAG.....TTTGGCCACGCGCAATATAT 1526
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
839 snAspIleGluLeuAspAspThrArgPheLeu.....GlnIuTr 852
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1527 CCGTATGTCATGTCAGATAGTTTACTGATCTTTAAACAAAAGAGGCA 1576
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
853 Asp.....IleSerAsnAlaIleProAspIleValTrpGluGlyValAs 867
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1577 TTCAGAGACCTTCAAGAGATGAGATGTCAATTGCTGAAGTGAAGGCC 1626
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
867 nThrLysThrLeuAspLysMet.....GlnAspAlaServ 879
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1627 ATACTAGACACTTGACACAGAAAAAGAAATGAGACAGATTACAAAGTA 1676
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
879 aLAspArgMetLeuGlnAsnGly..... 886
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1677 GAGCCACACAGGCGCTTGGACTCCAGTACTCAAGACAGGCTCATAGCGCT 1726
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
887 .....IleAsnLysGlnSerArgPheLeuAlaAs 896
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

```

```

1727 GAAAGCACTAACAAATTTGGCGTTCAGTTGAAGTTTTCACITCTGAG 1776
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
896 nlys.....AspleuglyLeuthrProlyMet.....AsnglnA 908
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1777 AAGCTGAATATATTCAG.....AAGAACTTGATGAGACACCAAGATTG 1820
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
908 snlleThleuileglnGlnleargHsileCySHslyslSerleu 924
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1821 CTCAGGGAATCTCAG.....GAAGCCCAAGATGAAGCT 1853
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
925 lleargMetleuGlnSerProleuSerAlaGlnSerArg 938
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

seq_name: p1r2:S68142

seq_documentation_block:
probable transcription factor SNAP - human
N/Alternate names: skeletal muscle abundant protein
C/Species: Homo sapiens (man)
C/Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 20-Sep-1999
C/Accession: S68142
R/Madsen, M.S.; Munk Petersen, C.; Gilleman, J.; Madsen, P.
Biochim. Biophys. Acta 1306, 14-16, 1996
A/Title: Cloning and sequencing of a human cDNA encoding a putative transcription factor
A/Reference number: S68142; WUID:96201699
A/Accession: S68142
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-757 <NIE>
A/Cross-references: EMBL:X87613; NID:g1009425; PIDN:CAA60949.1; PID:g1009426
C/Superfamily: unassigned bromodomain proteins; bromodomain homology
F:626-681/Domain: bromodomain homology <BR02>

alignment_scores:
      Quality: 198.00      Length: 363
      Ratio: 1.048      Gaps: 20
      Percent Similarity: 52.066      Percent Identity: 26.171

alignment_block:
US-09-687-230-1 x S68142 ..

Align seg 1/1 to: S68142 from: 1 to: 757

      8 GGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 57
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
421 G1yAlValProAlaThrSerMetGluProGluLeuArg..... 434
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
58 CTCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 107
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
435 .....SerGlnspleuAspGluGlnLeuGlySerThrAlaIag 448
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
108 GCGCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGG 157
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
448 lYcIuIlleuGlnAlaAspValAla.....Ile 457
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
158 GGACATGGGCAAGACCAAGAACCAAGTCCGACAAACCTCTACG 207
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
458 G1ylysglyAspGlu.....ThrProLeuTh 466
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
208 AGAGATATGTAGAGAGCCCTTGAAGCTGTCTCAAAAGTAGAGGAAC 257
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
466 rAsnVal.....LysThrGlnAlaSerProGluSerMetLeuSerP 480
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
258 GAAGTCACCAACTCTCCAGCGGCAAGCTCGGGCAAGACTCCAGCTCTT 307
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
480 roSerHsIsglySer.....AsnProIle 487
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
308 CGAAGACAAAAAGATCATGAC...AAACACAGAGACAGAAAGCGAAAA 354
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
488 .GluAspProleuGlnAlaGlnThrGlnHsLysPheGluMetSerAsps 504
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
355 AGAGAAAGAAAAAGAGAG.....AAGCAGATT 380

```

```

504 erleuLysGluGluSerGlyThrIlePheGlySerGlnIleLysAspAla 520
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
381 CCAAGG.....GAAGAAAAGGAGAAAAGCGAAGAGATTAAAGA 421
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
521 ProGluGlnAspGluGlnGluAspGlyValSerGlnAlaIaIaSerleuG 537
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
422 CGATTAAGAAAGACGACATCGA.....GACCGGTGGAGA 456
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
537 uGluProGlyGlnGluAspGlnGlyGlyTyrlLeuSerGlnMetAspA 554
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
457 ATGAG.....GCAGAAAAGAT..... 473
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
554 snGluProProValSerGluSerAspAspGlyPheSerIleHsAsnAla 570
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
474 ..CTCCAGTGTACAGCCCGCTGTGAGATTGACTTGCCTCGAGAAAGCC 520
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
571 ThrleuGlnSerHis.....ThrleuAlaAspSerIleProSerSerP 585
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
521 TCTCACAAGCTCTTACCC.....AAACAAAGAAAGTAGAA. 557
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
585 oAlaSerSerGlnPheSerValCysSerGluAspGlnGlnAlaIleGlnA 602
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
558 .CAGACACCCCTTCAGAAAGCTTTGATCAACGATGACAGACATTGCG 605
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
602 lGlnLysIleThrPlyLysAlaIleMetLeuValIlePargAlaAlaAla 618
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
606 AGAAAGATCCAAAGTCTTCTTTCATTCCTGTCGATGATTATATTCG 655
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
619 AsnHsIargTyrlAlaAsnValPheleuGlnProValHsAspIleAl 655
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
656 TCCGCTACTCCATGATCATTTAAACACCAATGATTTAGTACATGA 705
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
635 aProGlyTyrlHsSerIleValGlnArgProMetAspLeuSerThrIle 652
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
706 AAGAAAGATCAAGACATGACTATGATCCATAGAAAGACTAAAGAT 755
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
652 yslYsaHsIleGlnAsnGlyleuIleArgSerThrAlaGluPheGlnArg 668
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
756 AACTTCAACTAATGTGTACTAATGCCATGATTACAAATAACCCAGAG.. 803
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
669 AspIleMetLeuMetPheGlnAsnAlaValIleTyrlAsnSerSerAspH 685
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
804 .ACCATTATATTAACCTGCA.....AAGACCTGTGCATCTCAG 843
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
685 sAspValTyrlHsMetAlaValGluMetGlnArgAspValleuGlnI 702
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
844 GAATGAATAATCTTAGCCAGAGAAAGATTCAGAGCCGTGAAGCAG..... 887
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
702 leGlnGlnPheleuAlaThrGlnleuIleMetGlnHsSerGluSerGly 718
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
888 ..AGCATGACTTCATGCTGACTGACTGCAGAAAACTGAAAGCAGAAAGA 934
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
719 IleSerAlaLysSerleuArgGlyArgAspSerThrArgLysGln..... 733
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
935 TGGACAGACACCTCAGAGAGTGGGAGAGACGAGGCC 971
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
734 .....AspAlaSerGlu.....LysAspGlyGly 741

```

THIS PAGE BLANK (USPTO)

51 AsplysasnAspHisAspLysHisLysAspArgLysArgLysLysArgLys 67
 362 GAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 411
 67 slysglyglulysglulieProgluluglulysglulysArgLysArgArg 84
 412 GAGTTAAGGAGGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 461
 84 rgVallysglulnAspLysLysArgAspArgAspArgValgluAsnGlu 100
 462 GCAGAAAAGATCCAGTGTCCAGCCCTGTGAGATTAGACTGGCTGCC 511
 101 AlaGluLysAspLeuGlnCysHisAlaProValArgLeuAspLeuPro 117
 512 TGAGAGGCTCTCACAAGCTTTAGCCAAACAGAGAAGTGAACAGA 561
 117 ogulysProLeuThrSerSerLeuAlaLysGlnGluValGluGln 134
 562 CAGCCCTTCAGAGGCTTTGAATCAATGATGAGACAAATTCAGAGAAA 611
 134 hPProLeuGlnGluAlaLeuAsnGlnLeuMetArgGlnLeuGlnArgLys 150
 612 GATCCAGAGCTTTCTTTCATCTCTGTGAGTATTTATTCCTCTCG 661
 151 AspProSerAlaPhePheSerPheProValThrAspPheLeuAlaProG 167
 662 CTACTCCATGATCATTAACACCAATGATTTTATACCATGAAGAAA 711
 167 yTrSerMetIleIleLysHisProMetAspPheSerThrMetLysGlu 184
 712 AGATCAAGAACAAATGACTATCAGTCCATGAAGAACTAAAGCAATTC 761
 184 yLleLysAsnAsnAspTrpGlnSerIleGluLulnLeuLysAspAsnPhe 200
 762 AAATCATATGCTACTATGCGATGATTTCAATAAACAGAGGCCATTA 811
 201 LysLeuMetCysThrAsnAlaMetIleTrpAsnLysProGluThrIleTy 217
 812 TTATAAGGTGCAGAAAGAGCTGTTCAGCTCAGAGATGAATAATTCAT 861
 217 rTyTrLysAlaAlaLysLysLeuLeuHisSerGlyMetLysIleLeuSerG 234
 862 AGGAAAGAAATTCAGAGCTGAGACAGAGCATGACTTCATGGCTGACT 911
 234 InGluArgIleGlnSerLeuLysGlnSerIleAspPheMetAlaAspLeu 250
 912 CAGAAAACGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 961
 251 GlnLysThrArgLysGlnLysAspGlyThrAspTrpSerGlnSerIleG 267
 962 GAGAGGAGGCTGTGCGAGAGAGAGAGAGAGACTCTGGAGATGCCGAG 1011
 267 uAspGlyGlyCysTrpGlnArgGluArgLysAspSerGlyAspAlaGlu 284
 1012 CACAGCCCTTCAGAGCTCCAGCAAGGAGGAGGAGGAGGAGGAGGAG 1061
 284 lAhHisAlaPheLysSerProSerLysGlnAsnLysLysAspLysAsp 300
 1062 ATGCTTGAGATTAAGTTTAAAGCAATTAATTTAGAGAGAGAGAGAGCA 1111
 301 MetLeuGlnLysPheLysSerAsnAsnLeuGlnArgGlnGlnGln 317
 1112 GCTTGACCGCATCGTGAAGGAGATCGAGAGGAGGAGTGAAGCGGCTTG 1161
 317 nLeuAspArgIleValLysGlnSerGlyGlyLysLeuThrArgArgLeu 334
 1162 TGAACAGTCAAGTCAATTTGAAAGAGGAGGAGGAGGAGGAGGAGGAG 1211
 334 aLAsnSerGlnCysGlnPheGlnArgArgLysProAspGlyThrThr 350
 1212 TTGGAGCTTCATCCTGTGATCCATGTGAGAGAGGAGGAGGAGTACT 1261
 351 LeuGlyLeuLeuHisProValAspProIleValGlyLulnProGlyTrp 367

1262 CCTGTGAGACTGGAGATGACAACTGGAAGACTTCAGTCTGGAGTGAATA 1311
 367 sLeuValArgLeuGlnMetThrThrGlyArgLeuGlnSerGlyValAsnT 384
 1312 CTTTGACAGGGGCTTCAAAGAGATTAAGGAGCAACAACTCACTCAGTGT 1361
 384 hLeuGlnGlnLysPheLysGlnAspLysArgAsnLysValThrProValLeu 400
 1362 TATTGAAATTATGGGCGCTACAGTCTTATGCAGCCGATTAATGATCCAC 1411
 401 TyrLeuAsnTrpGlyProTrpSerSerTrpAlaProHisTyrAspSerTh 417
 1412 ATTTGCAAAATATCAGCAAGAGATGATTCGATTTAATTAATCAACCTAT 1461
 417 rPheAlaAsnIleSerLysAspAspSerAspLeuIleTyrSerThrTyrG 434
 1462 GGGAGACTCTGATCTTCACAGTGAATGATTTAGCATCCAGTATTTGGCC 1511
 434 LysLysAspSerAspLeuProSerAspPheSerIleHisGlnPheLeuAla 450
 1512 AGTGGCAAGATTATCCGTATGTCATGCGAGATGATGATGATGATTTT 1561
 451 ThrCysGlnAspTrpProTrpValMetAlaAspSerLeuLeuAspValle 467
 1562 AACAAAAGGAGGCGATTCAGAGCCCTACAGAGATGAGATGATGATGC 1611
 467 uThrLysGlyLysHisSerArgThrLeuGlnGlnLeuMetSerLeuP 484
 1612 CTGAAGATGAAGGCCATTAAGACACTGACACAGGAGGAGGAGGAGGAG 1661
 484 roGluAspGlnGlyHisThrArgThrLeuAspThrGlyLysGlnMetGlu 500
 1662 CAGATTACAGAAGTAGAGAGCCAGAGGCGTTTGACTCCAGTACTCAGA 1711
 501 GlnIleThrGluValGluProProGlyArgLeuAspSerSerThrGlnAs 517
 1712 CAGGCTCATAGCGCTGAAGAGCAGTAACAATTTGGCGTTCCAGTTGAAG 1761
 517 pArgLeuIleAlaLeuLysAlaValThrAsnPheGlyValProValGlu 534
 1762 TTTTTCAGCTGAGAGGAGTGAATATTCAGAGGAGGAGGAGGAGGAGGAG 1811
 534 aLysAspSerGlnGluAlaGluIlePheGlnLysLysLeuAspGluThr 550
 1812 ACCAGATTGCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1861
 551 ThrArgLeuLeuArgGluLeuGlnGlnAlaGlnAsnGlnArgLeuSerTh 567
 1862 CAGAGCCCTGGGAGACATGATCTCTCTGGGCTCCCTCATCAGAGAAT 1911
 567 rArgProProGlyAsnMetIleCysLeuLeuGlnProSerSerGluLysC 584
 1912 GCATCTTGCTGAAGCAG 1928
 584 yLleLeuLeuAsnLys 589

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-194-468-2
 seq documentation block:
 ; Sequence 2, Application US/08194468
 ; Patent No. 5750336
 ; GENERAL INFORMATION:
 ; APPLICANT: Montminy, Marc R.
 ; TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
 ; TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
 ; STREET: 444 South Flower Street, Suite 2000
 ; CITY: Los Angeles
 ; STATE: California


```

1392 yfrrgThrlYsAlaLeuPheAlaPheGluIleAspGlyValAspVal 1408
1410 ACATTTCGAATATATACAGCAAGATGATTCGATTATATC..... 1448
1409 CysPhePheGlyMetHisValGlnAspThrAlaLeuIleAlaProHisG1 1425
1449 .....TATTCACCTATGGGGAAGACTGTGATCTTCCAA 1482
1425 nileGlnGlyCysValTyrIleSerTyrLeuAsp..... 1436
1483 GTGATTTCAGCATTCATGATTTTGGCCACGTC..... 1517
1437 .....SerIleHisPhePheArgProArgCysLeuArgThrAlaVal 1450
1518 .....CAAGATTATCCGTA 1531
1451 TyrHisGluIleLeuIleGlyTyrLeuGluTyrValLysLysLeuValTy 1467
1532 TGTTCATGGCAGATAGTTTACTGGATGTTTACAAAGAGGG..... 1574
1467 rValThrAlaHisIleThrAlaCysProProSerGluGlyAspAspTyrI 1484
1575 .....CATTCAGAGCCCTA 1589
1484 IepheHisCysHisProProAspGlnLysIleProLysProLysArgLeu 1500
1590 CAAGAGATGAGATGTCATTCCTCGAAGATGAAGCCATCTAGACACAT 1639
1501 GlnGluTyrTyrLysLysMetLeuAspLysAlaPheAlaGluArgIleI 1517
1640 TGACACAGAAAAAGAAATGAGCAGATTCACAGATGAGCCACCAAGGC 1689
1517 eAsnAspTyrLysAspIle..... 1523
1690 GTTGGACTCCAGTACTCAAGACGCTCATAGCTGAAGCAGTAACA 1739
1524 ..PheLysGlnAlaAsnGlnAspArgLeuThrSerAlaLysGluLeuPro 1539
1740 AATTTTGGCGTTCACAGTGAAGTTTGAAGTGAAGCAGTAATATTT 1789
1540 TyrPhe.....GluGlyAspPheTr 1546
1790 CGAGAGAAACTGANTGAGACCAACGATGCTCAGGAGAACTCCAGGAAG 1839
1546 ProAsnValLeuGlnLysSer.....IleLysGluLeuGluGlnG 1560
1840 CCCAGATGAACT.....TTGACACACAGAGCC 1868
1560 IuGlnGluGlnLysLysGlnGlnSerThrAlaAlaSerGluThrPro 1576
1869 CCTGGACATGATCTGTCTTGGTCCCTCATGACAGAGAAATGCATCTT 1918
1577 GluGlySerGln.....GlyAspSer..... 1583
1919 GGTGAACAAGTACCAATATCTTAAAGATTCAGCAGCAATACTCCA 1968
1584 .....LysAsnAlaLysLysLysAsnAsnL 1592
1969 GGTGATATCGTAAGCAGTATGAGT...TCGAAAAGCAATGGGATTTCC 2015
1592 yslYThrAsnLysAsnLysSerSerIleSerArgAlaAsnLysLysLys 1608
2016 CATTCCTTCCCGCTCAGGAAACAACTT 2045
1609 ProSerMetProAsnValSerAsnAspLeu 1618
seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-227-536-2
seq_documentation_block:
; Sequence 2, Application US/08227536
; Patent No. 5658784
; GENERAL INFORMATION:

```

```

; APPLICANT: Eckner, Richard
; APPLICANT: Ewen, Mark
; TITLE OF INVENTION: NUCLEIC ACID, ENCODING TRANSCRIPTION
; TITLE OF INVENTION: FACTOR P300 AND USES OF P300
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESS: Weingarten, Schurgin, Gagnebin & Hayes
; STREET: Ten Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,536
; FILING DATE: 14-APR-1994
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Ph.D., Kathleen A.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: DFCI-308XX
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-2290
; TELEFAX: (617) 451-0313
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2414 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-227-536-2

alignment_scores:
    Quality: 216.00      Length: 740
    Ratio: 0.663         Gaps: 33
    Percent Similarity: 44.054    Percent Identity: 21.216

alignment_block:
US-09-687-230-1 x US-08-227-536-2 ..
Align seg 1/1 to: US-08-227-536-2 from: 1 to: 2414

17 CGCGCCGGC.....CGCGCGCCCGCCCTGCTCGCGCGCGG 54
|||||
870 ProProGlyProGlnSerGlnAlaLeuHisProProPro..... 882
55 GGCTCGCGGGCGCGCTCCGCCCTCGCGCTCGCGCGCGGAGCA 104
|||||
883 ....ArgGlnThrProThrProThrThrGlnLeuProGlnGlnValG 898
105 GCGCGCCCGCCAGCGCGCTGCGCGCGGGGCGCGGAGCGCGGCCG 154
|||
898 InProSerLeuProAlaAlaProSerAlaAspGlnProGlnGlnPro 914
155 GTGCGACATGGCGAGCAGCAAGAGCAAGATC..... 190
|||||
915 .....ArgSerGlnGlnSerThrAlaAlaSerValProThrProAsn 929
191 .....GACAAACACCTCTACGA.....GGAGTAT 215
929 aProLeuLeuProProGlnProAlaThrProLeuSerGlnProAlaValS 946
216 GTAGAGAAAGCCCTTGAAGCTGCTCAAGTAAAGAGAGAGAGCAAGTCAC 265
|||||
946 erIleGlnGlnGlnValLysAsnProProSer.....ThrSerSer 959
266 CGAACTCTTCACGGGCGAGCTCGGGGCGAGCATTCAGCCTCTTGAGAGACA 315

```

```

960 ThrGluValAsnSerGlnAlaIleAlaGluLysGlnPro..... 972
316 AAAAGATCATGACAAACAGAGAGAGAAAGCGGAAAGAAAGAAAGAA 365
973 .....SerGlnGluValLysMetGluAlaLysMetGluVal 984
366 GGAGAGAGACGATTCACAGGGGAGAGAGAGAGAGAGAGAGAGAGAG 415
984 AlaSpGlnProGluProAlaAspThrGlnProGluAspIleSerGluSer 1000
416 TAAGAGGATATAAAGAGCGAGATCGAGACCGGGTGGAGATAGCGAG 465
1001 LysValGlnAspCysLysMetGluSerThrGluThrGlu...GluArgS 1016
466 AAAAGATCTCCAGTGTACAGCCCTGTAGATTAGACTGGCTGCTGAG 515
1016 erThrGluLeuLysThrGluIleLysGluGluLysPcInProSerThr 1032
516 AAGCTCTCACAGCTCTTTAGCC.....AAACAGAGAGAGAGT 553
1033 SerAlaThrGlnSerProAlaProGluLysIleSerLysLysIlePhe 1049
554 AGAAGAGACACCCCTTCAGAGAGCTTTGATCACTGATGAGAGACAATTG 603
1049 elysProGluGluLeuArgGlnAlaLeuMetCProThrLeuGluAlaLeuT 1066
604 AGAGAGAGATCCAGATGCTTTC...TTTTCATTTCTGTGACTGATTTT 650
1066 yArgGlnAspProGluSerLeuProPheArgGlnProValAspProIn 1082
651 ATT.....GCTCTGCTACTCCATGATCATTAACACCCAGATGATTT 694
1083 LeuLeuGlyIleProAspLysPheAspIleValLysSerProMetAspLe 1099
695 TAGTACATGAAGAGAGAGATCAAGAGACATGACTATGATCCATAGAG 744
1099 uSerThrIleLysArgLysLeuAspThrGlyGlnIleGlnIleProTrpG 1116
745 AACTAAAGATTAATCAACTAATGTTACTAATGTCATGATTTACAT 794
1116 InTyValAspAspIleThrLeuMetPheAsnAsnAlaIleThrLeuLys 1132
795 AAACAGAGACCATTTATTAAGCTCCAAAGAGCTGTTGCACCTGAG 844
1133 ArgLysThrSerArgValIleLysIleLysSerLysLeu..... 1145
845 AATGAAATTTCTTACCGAGAGAAATTCAGAGCCTGAGAGAGAGATAG 894
1146 SerGluValIlePheGlnGlnIle...IleAspProValMetGlnSerLeu 1161
895 ACTTCATGGCT..... 905
1161 LysTyCysGlyLysLeuGlnLysPheSerProGlnThrLeuLysCys 1177
906 ...GACTTGCAGAAACTCGAAAGCAGAAAGATGAGACAGACCTCA 952
1178 TyrGlyLysGlnLeuLysThrIleProArgAspAlaThrIleTyrSer 1194
953 GAGTGGAGAGAGAGAGAGCTGC.....TGGCAGAGAG 984
1194 rGlnAsnArgLysIlePheCysGlnLysCysPheAsnGluIleGlnGly 1211
985 AGAGAGAGAGCTGTGAGATGCCGAGAGACAGCGCTTCAGAGATCCAGC 1034
1211 LysSerValSerLeuGlnLysAspProSerGlnProGlnThrThrIleAsn 1227
1035 AAAGAAAT...AAAAGAGAGAGAGAGATATGCTGTAAGATTAAGTT 1079
1228 LysGlnGlnPheSerLysArgLysAsnAspThrLeuAspProGluLeuP 1244
1080 ...AAAAGCAATATTTAGAGAGAGAGAGAGAGCTT..... 1115
1244 eValGluCysThrGluCysGlyArgLysMetHisGlnIleCysValLeuH 1261
1116 .....GAGCCGATGGTGAG 1130
1261 LSHSGluIleIleThrProAlaGluPheValCysAspGlyCysLeuLys 1277
1131 GAATCTGCA..... 1139
1278 LysSerAlaArgThrArgLysGluAsnLysPheSerAlaLysArgLeuP 1294
1140 .....GAAAGCTGACAGCGGCTGTGTGACAGTACAGTACGCG 1176
1294 oSerThrArgLeuGlnIleThrPheLeuGluAsnArgValAsnAspPheLeu 1311
1177 AATTGAGAGAGAGAGAGAGAGATGAGACAGAGCTGGAGCTTCAT 1226
1311 rGArgGlnAsnHisProGluSerIleGluValIleThrValArgValHis 1327
1227 CCGTGCATCCCATTTGAGAGAGCCAGGCTTACCTGCTGAGAGCTGCG 1276
1328 AlaSerAspLysThrValGlu.....ValLysProGln 1338
1277 AATGAGAACTGAGAGACTTCAGTGTGAGATGATTTGAGAGGCTTCA 1326
1338 yMetLysAlaArgPheValAspSerGly...GluMetAlaGlnSerPhe 1353
1327 AAGAGATTAAGAGAGAGAGAGAGTCACTCCAGTGTATATTGAATTAGG 1376
1353 ..... 1353
1377 CCGTACAGTCTTATGCACCGCATTTAGTCCATTTGCAATATACG 1426
1354 ProTyrArgThrLysAla.....LeuPheAlaPheGlnGlu 1365
1427 CAAGATGATTCGATTAATCAT.....TCAACCATAG 1461
1365 uIleAspLeuValAspLysCysPhePheGlyMetHisValGlnIleTyr 1392
1462 GGAAGACTCTGATCTTCAAGAT..... 1487
1382 LysSerAspCysProProProAsnGlnArgValIleLysSerTyrLeu 1398
1488 TTCAGCATCCATGATTTTGGCCAGCTGC.....CAAGTTATCC 1528
1399 AspSerValHisPhePheArgProLysCysLeuArgThrAlaValTyrH 1415
1529 GTATGATGAGCAGATAGTTACTGGATGTTTAAAGAGAGAGCAT 1578
1415 sGluIleLeuIle...GlyTyrLeuGluTyrValLysLysLeuGlyTyr 1431
1579 CCAGAGCCCTACAGAGATGAGATGTCATTCCTGAAAGATGAAGCCAT 1628
1431 hr.....ThrGlyHis 1434
1629 ACTGAGACACTTGACACAGAGAGAGAGAGATTCACAGAGATGAGA 1678
1435 IleThrAlaCysProProSerGluLysAspAspLysIlePheHisCysH 1451
1679 GCCACAGAGGCT..... 1691
1451 sProProAspGlnLysIleProLysProLysArgLeuGlnGluTyrTyr 1468
1692 .....TTGGAC..... 1697
1468 yLysMetLeuAspLysAlaValSerGluArgIleValHisAspTyrLys 1484
1698 .....TCCAGTACTCAAGACAGGCTCATACCGCTGAAAGCAGT 1735
1485 AspIlePheLysGlnAlaThrGluAspArgLeuThrSerAlaLysGluLe 1501
1736 AACAAATTTGGCTTCACAGTGAAGTTTGAATCCTGGAAGAGCTGAA 1785
1501 uProTyrPhe.....GluGlyAsp 1508

```



```

1178 TyrGlyLysGlnLeuCysThrIleProArgSpalathrTyrTyrSerTy 1194
953 GAGTGGGAGGAGCGAGGCTGC.....TGCGAGAGAG 984
1194 rGlnAsnArgTyrHisPheCysGlnLysCysPheAsnGlnIleGlnGly 1211
985 AGAAGAGAGAGACTCTGGAGATGCCGAGACACAGCGCTTCAAGAGTCCGACG 1034
1211 IuSerValSerLeuGlnLysAspProSerGlnProGlnThrThrIleAsn 1227
1035 AAGAAAT.....AAAAGAAAGACAAAGATATGCTTGAAGATAGTT.. 1079
1228 LysGlnGlnPheSerLysArgLysAsnAspThrLeuAspProGlnLeuPh 1244
1080 ....AAAAGCAATATATTAGAGAGAGACAGACAGCTT..... 1115
1244 eValGlnCysThrGlnCysGlnArgLysMetHisGlnIleCysValLeuH 1261
1116 .....GACCGCATGCTGAAG 1130
1261 LSHISGlnIleIleTrrProAlaGlnPheValCysAspGlnCysLeuLys 1277
1131 GAATCTGGA..... 1139
1278 LysSerAlaArgTyrArgLysGlnAsnLysPheSerAlaLysArgLeuPr 1294
1140 .....GGAAAGCTGACCGAGCGGCTTGTGAACAGTACGTGCG 1176
1294 oSerThrArgLeuGlnLysThrPheLeuGlnAsnArgValAsnAspPheLeuA 1311
1177 AATTGAAAGAAAGAAACCAAGATGGAACAGAGCTTGGAGACTTCCAT 1226
1311 rGArgGlnAsnHisProGlnSerGlnLysValThrValArgValAlaHis 1327
1227 CCTGTGATCCCATTTGTAGAGAGCGCAGCTACTGCTGTAGACTGGG 1276
1328 AlAserAspLysThrValGlu.....ValLysProGln 1338
1277 AATGACAACCTGAGAGACTTCACTGTGAGTGAATCTTGCAGGGGTCA 1326
1338 YMetLysAlaArgPheValAspSerGly...GluMetAlaGlnSerPhe. 1353
1327 AAGAGATAAAGAAAGACAAAGTCACTCCAGTGAATTATTTGAATTATGGG 1376
1353 ..... 1353
1377 CCTACAGTCTTATGCACCCGATATGATGCACATTTGCAAAATATCAG 1426
1354 ProTyrArgThrLysAla.....LeuPheAlaPheGlnGln 1365
1427 CAAGGATGATTCGATTTAATCTAT.....TCAAACCTATG 1461
1365 uIleAspLysValAspLeuLysPheGlnMetHisValGlnGlnLysArg 1382
1462 GCGAAGACTCGATCTCCAGCTGAT..... 1487
1382 LysSerAspCysProProProAsnGlnAlaGlnValTyrIleSerTyrLeu 1398
1488 TTCAGCATCCATGAGTTTGGCCACGTCG.....CAAGATTATCC 1528
1399 AspSerValHisPhePheArgProLysCysLeuArgThrAlaValTyrHis 1415
1529 GTATGTCAATGGCAGATGATTACTGAGTGTTTAACAAGAGGGGCAAT 1578
1415 sGlnIleLeuLeu...GlyTyrLeuGlnTyrValLysLysLeuGlnTyrT 1431
1579 CCAGGACCCCTAAGAGATGAGATGATGATGCCGGAAGATGACAGGCAAT 1628
1431 hr.....ThrGlnHis 1434
1629 ACTAGACACTTGCACACGAAAGAAATGAGACAGATTTACAGAGATAGA 1678
1435 IleTrrAlaCysProProSerGlnGlnLysAspArgTyrIlePheHisCysHis 1451

```

```

1679 GCCACACAGGCGCT..... 1691
1451 sProProAspGlnLysIleProLysProLysArgLeuGlnGlnTrrPyrL 1468
1692 .....TTGGAC..... 1697
1468 YLysMetLeuAspLysAlaValSerGlnArgIleValHisAspTyrLys 1484
1698 .....TCCAGTACTCAAGACAGCGCTCATAGCGCTGAAGACAGT 1735
1485 AspIlePheLysGlnAlaThrGlnAspArgLeuThrSerAlaLysGlnLe 1501
1736 AACAAATTTGGCGTCCAGTTGAAGTTTTCAGCTTGAGAGACCTGAAA 1785
1501 uProTyrPhe.....GlnLysAsp 1508
1786 TATTCACAGAAACAACTGTAGTACACACACAGATGCTCAGGAGCACTCAG 1835
1508 hettProAsnValLeuGlnGlnUser.....IleLysGlnLeuGln 1521
1836 GAAGCCCGAATGAACGT 1853
1522 GlnGlnGlnGlnGlnArg 1527

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-188-582-14

seq_documentation_block:
; Sequence 14, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comali, Lucio
; APPLICANT: Dynalac, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLHR, HOBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,582
; FILING DATE: 28-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1872 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

```



```

seq.name: /cgn2_6/prodata/2/1aa/5A_COMB.pep:US-08-646-715-14
seq_documentation_block:
  Sequence 14, Application US/08646715
  Patent No. 5637686
  GENERAL INFORMATION:
    APPLICANT: Tjian, Robert
    APPLICANT: Comali, Lucio
    APPLICANT: Dynalco, Brian D.
    APPLICANT: Hoey, Timothy
    APPLICANT: Ruppert, Siegfried
    APPLICANT: Tanese, Naoko
    APPLICANT: Wang, Edith
    APPLICANT: Weinzierl, Robert O.J.
  TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS, NUCLEIC ACIDS ENCODING TAIS AND METHODS OF USE
  NUMBER OF SEQUENCES: 36
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: FLEHN, HONBACH, TEST, ALBRITTON & HERBERT
    STREET: 4 Embarcadero Center, Suite 3400
    CITY: San Francisco
    STATE: California
    COUNTRY: USA
    ZIP: 94111-4187
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/646,715
    FILING DATE: 09-MAY-1996
    CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/188,582
    FILING DATE: 28-JAN-1994
    ATTORNEY/AGENT INFORMATION:
      NAME: Osman, Richard A
      REGISTRATION NUMBER: 36,627
    REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
    TELECOMMUNICATION INFORMATION:
      TELEPHONE: (415) 781-1989
      TELEFAX: (415) 398-3249
      TELEX: 910 277299
  INFORMATION FOR SEQ ID NO: 14:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 1872 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
    MOLECULE TYPE: peptide
    US-08-646-715-14

alignment_scores:
  Quality: 181.50          length: 561
  Ratio: 0.717            Gaps: 21
  Percent Similarity: 45.098   Percent Identity: 21.034

alignment_block:
  US-09-687-230-1 x US-08-646-715-14  ..
Align seg 1/1 to: US-08-646-715-14 from: 1 to: 1872

177 AGAAGCAACAATCGGACAAACCTTACGAGAGTAGTAGAGAACC 226
|||||.....:::|||||:
1354 LYSLSAGTGATGVALGLYThrThrValHisCysAspTyrLeuasnArgTr 1370
227 CTGGAAGCTGTCTCAAAGTAAGSAGAACGAAGTCACCGAACTGCCA 276
|-|||:::|||||:

```

[illegible]


```

1488  ....
1489  rGlnIleSerGlnSerMetLeuAspLeuCysAspGluLys..... 1501
529  GCTCTTAGCAACAAGAGATAGACAGACACCCCTTCAAGAGCT 578
1502  ....
579  TTGAATCAACATGATGAGACAAATTGCAGAGAAAAGATCCAAAGCTTTC 626
1516  IleAsnProLeu.....LeuAspAspAspGlnValAlaPheSe 1529
626  ....
1529  rPheIleLeuAspAsnIleValThrGlnLysMetMetAlaValProAsps 1546
627  ....
1546  eRTPrProPheHisProValAsnLysLysPheValProAspLysTyr 1562
669  ATGATCATTTAAACACCCCAATGATTTTATGACATGAAGAAAAGATCAA 718
1563  LysValIleValAsnProMetAspLeuGluThrIleArgLysAsnIleSe 1579
719  GAACATGATATCATGCTCCATAGACAACTAAAGATTAAGTCAACTAA 768
1579  rLysHisLysTyrGlnSerArgGlnSerPheLeuAspAspValAsnLeu 1596
769  TGTGTCTAATGCCATGATTTTAAATTAACAGACACCATTTATTTAA 818
1596  IeLeuAlaAsnSerValLysTyrAsnGlnProGlnSerGlnTyrThrLys 1612
819  GGTGCAAGAAGCTGTGTCACACGACGAAGAAAATTTCTTAGCCAG.... 863
1613  ThrAlaGlnGlnIleValAlaAsnValCysTyrGlnThrLeuThrGluTyrAs 1629
864  .GAAGATTCAGAGCCTGAG..... 884
1629  pGluHisLeuThrGlnLeuGlnLysAspIleCysThrAlaLysGluAla 1646
885  ....
1646  IeLeuGlnGluAlaGlnLeuGlnSerLeuAspProMetThrProGlnPro 1662
915  AAACCTGGAAGCAGAAAGATGCAACAGACACCTCACAGAGT..... 956
1663  TyrThrProGlnProProAspLeuTyrAspThrAsnThrSerLeuSerMe 1679
957  .GGGAGGAGGAGGCGTGGCAGAGAGAGAGAGAGAGAGCTCT.....G 999
1679  tSerArgAspAlaSerValPheGlnAspGlnSerAsnMetSerValLeuA 1696
1000  GAGATGCCGGAAGACAGACGCTTCAAGAGTCCAGCAAGAAAATTAAG 1049
1696  spIleProSerAlaThrProGlnLysGlnValThrGlnGlnGluLysP 1712
1050  AAAGACAAAGAT..... 1061
1713  GlysAspLysPheLeuAlaAspGlnGluGlnGlyThrValGlnGlnPro 1729
1062  ....
1729  nAlaSerValLeuTyrGlnAspLeuMetSer.....GluLys 1743
1102  AGCAGAGACAGCTTGAACCGCATGCTGAAGGAATCTGGAGGAAAGCTG 1151
1743  LysAspAspGlnGluAspAlaGlySerAspGlnGluLysAspAsn.... 1757
1152  AGCGGCGCTTGTGAACACTCAGTGGCAATTTGAAGAAGAAAACCAAGT 1201
1758  ....
1202  AACACAGACGTTGGAGCTTCCTCCATCTGTGATCCCATTTGTAGAGAG 1251

```

```

1771  rAspValGlySerGlyIleArgProLysGlnProArgMet..... 1785
1252  CAGGCTACTGCTGCTGATGAGACTGGGAATGACAACTGAGACCTTCA 1301
1786  ....
1302  GGAGTGAATACCTTTGCCAGGGGCTTCAAGAGAGATTAAGACAAAGT 1351
1789  ....
1352  .TCCAGTGTATAT..... 1364
1799  rMetMetSerTyrGlnGlyAspGlyGlnAlaSerHisGlyLeuGlnA 1816
1365  ....
1816  spSerAsnIleSerTyrGlySerTyrGlnGluLysProAspProLysSer 1832
1401  ...TATGCTCCACATTTGCCAAATATC..... 1424
1833  ThrGlnAspThrSerPheSerIleGlyGlyTyrGlnValSerGln 1849
1424  .... 1424
1849  uGlnGluAspGlnGluGlnGlnGlnArgSerGlyProSerValLeuS 1866
1425  ....
1866  eArgGlnValHisLeuSerGlnAspGlnGlnAspSerGlnAspPheHis 1882
1455  ACCTATGGGAGAGACTGTGATCTTCCAAAGTCAAT 1487
1883  IleAlaGly...AspSerAspLeuAspSerAsp 1892

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-646-715-11

seq_documentation_block:
; Sequence 11, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,715
; FILING DATE: 09-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/188,582
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627

```



```

883 GTTGCACCTAGAGATGAAAAATTTCTTAGCCAGAGAAATTCAGAGCCCTGA 882
      :|||:|||||:
571 .....AsnGlnGlyCysProArgSerLeu 579
      :|||:|||||:
883 AGCAGAGCATAGACTCATGGCTGACTGCAGAAACTCGAAAGCAGAAA 932
      :|||:|||||:
579 LUGLUSPLeuGlnThrLeuLysSerLeuGlnLysGlnLysAsnGlyAla 595
      :|||:|||||:
933 ...GATGGAACAGACACTCAGAGAGTGGGAGAGCGAGGCTGCTGCA 979
      :|||:|||||:
596 IleLysGlyCysGlyLysSerGlnThrSerArgLysArgLysCys... 610
      :|||:|||||:
980 GAGAGAGAGAGAGACTCTGGAGATGCCAAGACACCGCTTCAGAGATC 1029
      :|||:|||||:
611 ArgGlnLeuLysProThrGlyLysGlnLysPThrGlnThrLeuGlnSerL 627
      :|||:|||||:
1030 CCAGCAAAAGAAATATAAAGAAAGCAAGATATGCTTGAAGTATGTT 1079
      :|||:|||||:
627 eugGlnLysGlnLysGlnLysLeuMetLysSerLeu... 638
      :|||:|||||:
1080 AAAAGCAATATTTTAGAG...AGAGACGAGAGCA 1111
      :|||:|||||:
639 ...GlnGlyAsnLeuGlnThrPheLeuPheProGlyThrGlnAsnGln 654
      :|||:|||||:
1112 GCTTACCCGATCTGAGAGATCTGGAGAAAGCTGACCGCGGCTTG 1161
      :|||:|||||:
654 MetValSerSerLeuGlnGlnLysLeuLysSerLeuThr... 667
      :|||:|||||:
1162 TGAAACATCAGTCGCAATTTGAAAGAGAAACAGATGAAACAGAGAG 1211
      :|||:|||||:
668 .....AlaLeuGlnLysGlnLysGlnGlnLysProLeuArgSerProGlu 681
      :|||:|||||:
1212 TTGGAGCTTCTCCATCTGTGGATCCCATTTGAGAGAGCCAGGCTACG 1261
      :|||:|||||:
682 ValGlyAspGlnGlnAlaLeuArgProLeuThrLysGlu... 694
      :|||:|||||:
1262 CCGTGTAGACAGCGGAATGACAACTGGAAGACTTCAGTCTGAGTGAATA 1311
      :|||:|||||:
694 ..... 694
      :|||:|||||:
1312 CTTTGCAGGGGTTCAAGAGATAAAGACAAAGTACTCCAGTGTTA 1361
      :|||:|||||:
694 ..... 694
      :|||:|||||:
1362 TATTGAAATATGGCCCTACAGTTCTTATGACCGCATTATGACTCCAC 1411
      :|||:|||||:
695 .....AsnGlnGlnProLeuArgSerLeuGlnLysGlnLysGlnLys 709
      :|||:|||||:
1412 ATTTCGAATATACAGCAAGATGATTCGATTATTCATTCACACTATG 1461
      :|||:|||||:
709 ArgPheArgSerLeuGlnLysGlnLysGlnLysProLeuLysThrLeuGln 726
      :|||:|||||:
1462 GGAAGAGCTGATCTT...CCAAAGTGAATTCAGATCAGATGTT 1505
      :|||:|||||:
726 LUGLUSPLeuGlnThrLeuValArgProLeuGlnThrGlnAsnHisLysSer 742
      :|||:|||||:
1506 TTGGCCAGCTGCCAAGATATCCGATATGTCATGCGAGATAGTTACTGA 1555
      :|||:|||||:
743 LeuArgSerLeuGlnGlnLysGln... 749
      :|||:|||||:
1556 TGTTTTACAAAGAGAGGAGCATTCAGACCCCTACAGAGATGAGATGT 1605
      :|||:|||||:
750 .....AspGlnGlnThrLeuArgThrLeuGln... 758
      :|||:|||||:
1606 CATTCGCTGAAGATGAGGCGCATATCAGACACTTACACAGAAAGAA 1655
      :|||:|||||:
759 ...LysGlnThrGlnGlnArgArgSerLeuGly...Glu 770
      :|||:|||||:
1656 ATGAGAGCATTTACAGAGTAGAGCCAGCGGCTTTGAGCTCCAGTAC 1705
      :|||:|||||:
771 GlnAspGlnMetThr...LeuArgProGlnLysValAsp... 783
      :|||:|||||:
1706 TCAAGACAGCGCTCATAGCGCTGAAGACAGTA...ACAATTTTGGCGTTC 1752

```

```

784 .....LeuGlnProLeuLysSerLeuAspGlnGlnLeuAlaArgP 797
      :|||:|||||:
1753 CACTGAGATTTTGTACTCGAAGACCTGAATATTCGAGAAACTT 1802
      :|||:|||||:
797 roLeuGln...AsnGlnAsnGlnLysLeuLysSerLeu 809
      :|||:|||||:
1803 GATGAG...ACCACGATTTGCTCAG 1825
      :|||:|||||:
810 LysGlnGlnLysValGlnAlaValLysSerLeuGlnThrGlnLeuGln 826
      :|||:|||||:
1826 GGAATTCAGAGAAAGCCAGAGATGACGTTTGAACACC...AGACCCCTTG 1872
      :|||:|||||:
826 uSerLeuLysSerAlaGlyGlnGlnLysLeuGlnThrLeuLysSerProG 843
      :|||:|||||:
1873 GGAACATGATCTCTCTGGGCGCTCATCAGAGAAATGACATCT 1917
      :|||:|||||:
843 LuthrGlnAlaProLeu...TriThrProGlnGlnLysLysSer 857
      :|||:|||||:
seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-687-080-51
seq_documentation_block:
; Sequence 51, Application US/08687080
; Patent No. 5963427
; GENERAL INFORMATION:
; APPLICANT: Gregory Doljanov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Denlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,080
; FILING DATE: 17-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/592,126
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TRANS. OF RAD50 cDNA (SEQ. 54), NT.
; INDIVIDUAL ISOLATE: 389 TO 4324
; US-08-687-080-51
alignment_scores:
Quality: 154.50 Length: 750
Ratio: 0.456 Gaps: 29
Percent Similarity: 45.200 Percent Identity: 17.867

```



```

1411 CATTGGCAATTAACACAGGATTCATCTCATTTAACTATTCACCACTT 1460
      :::: |||::::::::::||| :: |||
753 yaleuclnlnsalnsalngasplleclnrgyleuylsasnaspilleglu 769
      ::::: ||| :::::
770 glunglunthrileuclenlythrilemetproglungluciseralaly 786
      ::|||
1511 CACGTGCCAAGATTATCCGATTCTCATGCAGATAGTTACTTGATGTTT 1566
      ::||| ||| |||||
786 svalys.....Leuthraspalt 793
1561 TACCAAAGAGGCGATTCACGACCCTTACAAGATGAGATGCTCATTTG 1610
      ::: |||::::::::::||| ::::
793 hrilemetgluarphneglmecgluleuylsaspvalgluarglysile 809
1611 CCTGAAGATGAGGCCCATACTAGACACTTGACACAGAAAAGAATAATGA 1666
      ::::: ::::: ||| |
810 AlaglnglnAlalaLysleuglnglylleaspleuaspargthValGI 826
1661 GCAGATT..... 1667
      ::|||:
826 nglnvalaInglnlgulyslnglnlulysglnhlslsleuaspthrVals 843
1668 ..ACAGAGTAGACCCACACAGGGCGTTTGACTTCCTACTCAAGACAGG 1715
      ::::: ||| ::::: ||| :::::
843 eeserYslIegIuleuasnArglysleuIleGlinaSpInglnlgln 859
1716 CTCATAGCGGTGAAGACAGTACAAATTTTGGCGTCCAGTTGAAGTTTT 1765
      :|||: |||::::::::::|||
860 IlcglnhlsleuylserThrThraS..... 868
1766 TGACCTGTGAAGAAGCTGAAATATTTCCAGAAAGAACTTGATGAGACCA 1815
      |||::: ::::: |||:::
869 .....GluleuylserglulysleuglnIleSerThr 880
1816 GATTGCTCAGGCAGACCTCAC.. GAACCCAGATGGAACCTTGAGACAC 1862
      ::||| |||::: ||| ||| ::::: ||| ||| ||| ||| |||
880 snleuclnhrAgArvglnglnleuglnglunlrvalgluIeSerThr 896
seq_name: /cgn2-6/plodata/2/laa/5B.COMB.pep:US-08-593-126-148
seq documentation block:
Sequence 148: Application US/08592126
Patent No. 5821091
GENERAL INFORMATION:
APPLICANT: Gregory Doljanov
TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
PROTEIN OF INVENTION: Polypeptides
NUMBER OF SEQUENCES: 151
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,126
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0950

```

[illegible]

```

623 TTTCCTTTCATTTCCGTGACGATTTTATTTGCTCTGCTACTCCATGTA 672
    |||
406 .....GlyGluAlaLysT 410
673 TCATTAAACCCCAATGGATTTCATACCATGAAGAAGAAATCAAGAAC 722
    |||
410 hrlaasnGlnLeuMetAsnAspPheAlaGlnLysGlnThrLeuLysGln 426
723 AATGACATGATGTCATAGAAAGAACTAAAGCATTAACATTC...AAGCAATAT 769
    |||
427 LysGlnIleAspGlnIleArgAspLysLysThrGlyLeuLysArgIleI 443
770 GTGTACTAATGCATGATTTCATATAAACAGAG.....ACCA 807
    |||
443 eGluLeuLysSerGlnIleLeuSerLysGlnAsnGlnLeuLysAsnV 460
    |||
808 TTTATTTTAA.....GCTGCAAGAAAGCTGTGCAC 839
    |||
460 allYsTYrGlnLeuGlnGlnLeuGlnLysSerAspArgIleLeuGln 476
840 TCAGGAATGAAATTTCTTACCCAGCAAGAA..... 869
477 LeuAspGlnIleLeuIleLysAlaGlnArgGlnLeuSerLysAlaGln 493
    |||
870 .....ATTTCAGACCTGAAGCAGACGAT.....GACT 897
    |||
493 sAsnSerAsnValGlnThrLeuLysMetGlnValIleSerLeuGlnAsnG 510
898 TCATGCTGCTGCGAAGAACTCGAAGAACGAAAGATGGAACAGACACC 947
    |||
510 LuLysAlaAspLeuAspArgThrLeuArgLysLeuAspGlnLysMetGln 526
948 TCACAGAGTGGGAGCAGCAGCGCTGTCGACAGAGAG.....AGAGA 991
    |||
527 GlnLeuAsnHisThrThrThrArgThrGlnMetGlnLysLeuThrLys 543
992 GGACTCTGAGAGATGCGAAGCAGCAGCTTCAGAGTCCAGCAAGAA. 1040
    |||
543 sAspLysAlaAspLysAspGlnIleArgLysIleLysSerArgHis 560
1041 .....AATMAAGAAAGAAC 1055
560 eAspArgLeuThrSerLeuGlnLysTyrPheProAsnLysLysGlnLeu 576
1056 AAGATATGCTTGAAGATTAAGTTTAA..... 1082
    |||
577 GlnAspArgLeuHisSerLysSerLysGlnLysAsnGlnThrArgAspArg 593
1082 ..... 1082
593 gLeuAlaLysLeuAsnLysGlnLeuAlaSerSerGlnGlnAsnLysAsnH 610
1083 .....AGCAATTAATTTAGAGAGAGCAGAGCAGCT..... 1115
    |||
610 IsIleAsnAsnGlnLeuLysArgArgGlnGlnLysLeuSerSerTyrGln 626
1116 GACCGCATGCTGAAGAAATCTGAGAGAAAGCTGACAGCGGCTTGTA 1165
    |||
627 AspLysLeuPheAspValCysGlnSerGln.....As 637
1166 CAGTCAGTGCATTTGAAGAAAGAAAGCAAGT..... 1199
    |||
637 PheGlnLysSerAspLeuAspArgLeuLysGlnGlnLysSerL 654
1200 .....GGAACAGCAGCTTGGACTTCTCCATCT 1229
    |||
654 ySglnArgAlaMetLeuAlaGlyAlaThrAlaVal...TyrSerGlnPhe 669
1230 GTGATTCCTCATTTGAGAGAGCCAGCTACTGC.....CTGGTAG 1270
    |||
670 IleThrGlnLeuThrAspGlnAsnGlnSerCysCysProValCysGlnArg 686

```

```

1271 ACTGGAAATGACAACCTGGAAAGACTTCAGTCTGGAGTGAATACTTGCAG 1320
    |||
686 gValPheGlnThrGlnAlaGlnLeuGlnGlnValIleSerAspLeuGln 703
1321 GG..... 1322
    |||
703 eLysLysLeuArgLeuAlaProAspLysLeuLysSerThrArgLysGlnLeu 719
1323 ...TTCAAGAGGATTAAGAAAGCAAA.....GTCACTCCAGTGT 1360
    |||
720 LysLysLysGlnLysArgArgAspGlnMetLeuGlnLeuValPheMetArg 736
1361 ATATTGAATTAATGAGCCCTACAGTTCTTATGCACCGCATTAATGACCA 1410
    |||
736 gGlnSerIleIleAspLeuLysGlnLysGlnLysGlnLysProGlnLeuArgSnl 753
1411 CATTTGCAATATTCAGCAAGAGATTCGATTTAATCTAATTCACCTAT 1460
    |||
753 ySLeuGlnAsnValAsnArgAspIleGlnArgLeuLysAsnAspIleGln 769
1461 GGGGAGACTCTGATCTTCCAGTGAATTCAGCATTCACATGAGTTTGGC 1510
    |||
770 GlnGlnGlnThrLeuLeuGlnLysThrIleMetProGlnLysSerAlaLys 786
1511 CAGCTGCAGATATATCCGATATGTCATGCGAGATGATTACTGATGTT 1560
    |||
786 sValCys.....LeuThrAspValT 793
1561 TAACAAAGAGAGGAGCTTCAGACCCCTACAGAGATGAGATGTCATTC 1610
    |||
793 hrlMetGlnArgPheGlnMetGlnLeuLysAspValGlnArgLysIle 809
1611 CTTGAGATGAGGCCATTAAGACATTCGACACTTGACACAGAAAGAAATGCA 1660
    |||
810 AlaGlnAlaIleAlaLysLysGlnGlnLysLeuAspArgThrValGln 826
1661 GCAGATT..... 1667
    |||
826 nGlnValAsnGlnGlnLysGlnGlnLysGlnHisLysLeuAspThrValS 843
1668 ...ACAGAAATGAGCCACAGCGGCTTGGACTCCAGTACTCAAGACAG 1715
    |||
843 eSerLysIleGlnLeuAsnArgLysLeuIleGlnAspGlnGlnGln 859
1716 CTCATTAAGCGCTGAAGACAGTAACTAATTTGGCGTCCAGTTGAAGTTT 1765
    |||
860 IleGlnHisLeuLysSerThrThrAsn..... 868
1766 TGACTCTGAAGAGCTGAATATATCCAGAGAAAGCTTGATGAGACACCA 1815
    |||
869 .....GlnLeuLysSerGlnLysLeuGlnIleSerThr 880
1816 GATGCTCAGGAGACTCCAG...GAAGCCAGAAATGAAGCTTGGACAGC 1862
    |||
880 snLeuGlnArgArgGlnGlnGlnGlnGlnGlnGlnValGlnLeuSerThr 896

```

seq_name: /cgn2_6/protdata/2/1aa/5a_comb.pep:us-07-853-913-2

seq_documentation_block:

; Sequence 2, Application US/07853913

; Patent No. 5338839

; GENERAL INFORMATION:

; APPLICANT: McKay, Ronald D.G.

; TITLE OF INVENTION: Nestin Expression As An Indicator of

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS: Neuroepithelial Tumors

; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.

; STREET: Two Militia Drive

; CITY: Lexington

; STATE: Massachusetts

; COUNTRY: U.S.A.


```

966 GGAGGCTGCTGCACAGAGAGA..... 989
    |||
527 GlnGlyProLeuGlnLysGlnThrLeuLysAlaLeuGlyGlnProLe 543
990 .....GAGAGCTTGA..... 1001
543 uMetSerLeuLysIleGlnAsnTyrGlnThrAlaGlyLysGlnCysA 560
1002 .....GATGCCGAAACACAC.....GCCCTCAAGAGTCCACAGAAAGAA 1040
560 snSerSerThrGlnGlyHisLeuGlyThrLeuGlnGlyProGlnLysGln 576
1040 ..... 1040
577 LysGlnIleProLeuLysSerLeuGlnGlnLysAsnValGlnSerGlnL 593
1040 ..... 1040
593 sThrLeuGlnAsnGlyValProValLeuSerGlnLeuGlnLysGlnA 610
1041 ..AATAAAAGAAAGACAAAGATATGCTTGAAGATTAAG..... 1076
610 sPThrArgThrGlnAspGlnLysLeuMetSerProLysGlyThrLeuLys 626
1077 .....TTTAA 1081
627 ArgPheSerSerLeuGlyLysGlnSerGlnLysValAlaArgProSerL 643
1082 AAGCAATTAATTAGAG..... 1097
643 sGlnGlyLysAsnLeuGlnSerTrpThrAlaPheLysGlnGlnSerGlnHis 660
1098 .....AGAGACGAGAGACGCTTGAACCCGATGTCG... 1127
660 rLeuGlnGlyPheProGlyAlaGlnLysAspGlnMetLeuGlnArgLeuValGln 676
1127 ..... 1127
677 LysGlnLysAspGlnSerPheProArgSerProGlnGlnLysAspGlnAla 693
1128 .....AAGCAAT 1134
693 acYArgProLeuGlnLysGlnAsnGlnLysProLeuGlnLysTrpGlnLys 710
1135 CTGAGAGAAAGCTGACCGCGCTTGTGAACAGCTCAGTGGCAATTTGAA 1184
710 lAglnGlyGlnIleLeuGlnLysIleGlnLysGlnSerGlnLys 726
1185 AGAGAAAACCA..... 1196
727 LeuArgSerProGlnGlnLysAspGlnLysAlaGlyArgSerLeuGlnLys 743
1197 .GATGAGAAACGACGTTGGACTTCCATCCGTGGATCCCAT... 1241
743 sGlnAsnGlnGlnLysProLeuGlnLysTrpGlnLysAlaGlnAspGlnMetLeu 760
1241 ..... 1241
760 luaArgLeuIleGlnLysGlnSerGlnLysSerProGlnLys 776
1242 .....GTAGAGAGACCAAGCTGCTGCTGAGTGGATGATGAC 1282
777 AsnGlnArgIleGlyLysPro.....LeuGlnArgGlnAsnGlnLys 790
1283 AACT.....G 1287
790 sSerLeuArgTyrLeuGlnLysAsnGlnLysThrPheValProLeuGln 807
1288 GAAGACTTTCAGTCTGAGTGAATCTTTCAGAGGCTTCAAGAGAGATAA 1337
807 errArgAsnGlnArgProLeuArgSerLeuLys...ValGlnGlnGlnLys 822

```

```

1338 AGAACAAGAGTCACTCCAGTGTATTATTGAATTATGGCCCTACAGTTC 1387
    :|||
823 GlnArgIleValLysProLeuGlnLysValSerGlnAspSerLeuGlySe 839
1388 TTATGCAACCGCATATATGATCCACATTTGCCAAATATACAGCAGTAT 1436
839 rLeuAlaGlnGlnLysAsnValGlnProLeuArgTyrLeuGlnGlnLysAsp 856
1437 .....TCTGATTTAACTTATTCACCTATGGGAA 1466
856 ysiLeaAsnLysSerLeuGlnLysAspLysThrHisLysSerLeuGlySer 872
1467 GACTCTGATCTTCCAGTGAATTCAGCATC..... 1496
873 LeuGlnAspArgAsnGlyAspSerIleIleProGlnGlnSerGlnLys 889
1496 ..... 1496
889 rGlnValSerLeuArgProProGlnGlnLysAspGlnArgIleValAsnH 906
1497 .....CATGAGTTTTGGCCACGTCGCCAAGATTAATCCG 1529
906 lSLeuGlnLysGlnSerGlnGlnPheSerArgSerSerGlnGlnLys 922
1530 TATGTCATGGCAGATTAAGTCTGATGTTTAAACAAAGAGGCGCATTC 1579
923 GlnValMetGlnArgSerLeu.....GlnGlnGlnLysAsnH 934
1580 CAGAGCCCTACAA.....GAGATGGAGATGTCATTCGCCAGATGAC 1623
934 sGlnSerLeuSerSerValLysGlnLysAspGlnMetValGlnSerGlnL 951
1624 GCCATTAAGACACTTGACACAGAAAGAAATGACAGATTAACAGAA 1673
951 euGlnLysGlnSerGlnAspSerGlyLysSerLeuGlnLysAspGlnSerGln 967
1674 GTAGAGCCACAGGCGCTTGGACTCCAGTACTCAAGACAGCTCATAGC 1723
968 ...GlnThrPheGlyProLeuGlnLysGlnAsnAlaGlnSerLeuArgSe 983
1724 GCTGAAGACAGTAAACAATTTGGCGTTCAGTTGAAGTTTTCAGTCG 1773
983 rLeuAlaGlyGln.....AspGln 990
1774 AAGAAGCTGAATATTCCAGAGAAATTCATGATGACACACAGATGCTC 1823
990 lGlnLys.....GlnLysLeuGlnGlnLysThrGlnGlnLysThrLeu 1002
1824 AGGGAAGCTCCAGAGAGCCGAGATGAACGTTTGACACACAGAC..... 1866
1003 ArgAlaValGlyLysAsnGlnLysMetAlaValSerProProGlnLysValA 1019
1867 .....CCCGTGAACATGAT.....CTGTC 1887
1019 sProGlnLysProLysProLeuGlyAsnAspGlnGlnIleAlaArgSe 1035
1888 TCTTGGGTCCTCATCAGAGAA.....ATGCACTTCTGTGAACAAGTGA 1931
1035 rLeuGlnLysGlnAsnGlnLysSerLeuValSerLeuLysGlnLysGly 1052
1932 CCAATATATCTTAAGAAATTCAGACAGCAAGTAACTCCAGGTATATGTA 1981
1052 lGlnLysValLysSerLeuGlnLysThrGlnIleIleLeuProLeuGlnLys 1068
1982 GCACGTATGAGATTCCAAAGCAAGTGGGATTTCCATCTTCCCGCCGTC 2031
1069 AlaGlnGlnLysProLeuArgArgLysSerIleAspThrGlnLysProLe 1085
2032 ATGGAACA 2041
1085 uTrpSerThr 1088

```

```

seq_documentation block:
; Sequence 2, Application US/08663112
; Patent No. 5849503
; GENERAL INFORMATION:
; APPLICANT: WAKATSUMA, Masako
; APPLICANT: KURITA, No. 5849503iko
; TITLE OF INVENTION: MUTANT PROTEINS OF HUMAN DNA
; TITLE OF INVENTION: TOPOISOMERASE I
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Parabow, Garrett &
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/663,112
; FILING DATE: 26-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Elnaudi, Carolyn P.
; REGISTRATION NUMBER: 32,220
; REFERENCE/DOCKET NUMBER: 06609,1488-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4400
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 765 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-663-112-2

alignment_scores:
      Quality: 148.00      Length: 499
      Ratio: 0.695      Gaps: 23
Percent Similarity: 42.685      Percent Identity: 20.040

alignment_block:
US-09-687-230-1 x US-08-663-112-2 ..

Align seg 1/1 to: US-08-663-112-2 from: 1 to: 765

159 GACATGGGCGACAGACAGACCAAGTGGAC.....194
||| :|||:|||||:|||||:|||||:|||||:|||||:
30 AsparGluHisArgHisLysGluHisLysGluLysAsparGluLys 46
195 ....AAACACCTTACGAGGATGTAGAGAACCCCTTGAACTGGTCC 240
||||| :|||:|||||:|||||:|||||:|||||:
46 sSerLysHisSerAsnSerGluHisLysAspSerGluLysHisLys 63
241 TCAAAGTAGAGGAGAACGATCCGCACTCCACGGCAGCTGGGG 290
||| :|||:|||||:|||||:|||||:
63 LysLys.....GluLysThrLysHisLysAspLysSer... 74
291 CACGACTCCAGCCTCTCGAAGACAAAGATCATGACAAACAAGGA 340
||||| :|||:|||||:|||||:|||||:
75 .....GluLysHisLysAspLysHisLysAs 83
341 CAGAAACGGAAAAAGAAAGAAAGAGAAAGAGATTCAGGGGAG 390
||||| :|||:|||||:|||||:|||||:
83 parGAspLysGluLysArgLysGluLysValArgAlaSerGlyAsp 100

```

```

391 AAAAGGGGAAAAACGAGAGAA.....413
||| :|||:|||||:|||||:|||||:|||||:
100 IalysIleLysLysGluLysGluAsnGlyPheSerSerProProGluIle 116
414 .....GTTAGGAGGAT..425
||| :|||:|||||:|||||:|||||:|||||:
117 LysAspGluProGluAspAspLysTyrPheValProProLysGluAspIle 133
426 .....AAAAAGCGAGATCGAGACCGGGTGGAG.....455
||||| :|||:|||||:|||||:|||||:|||||:
133 eLysProLeuLysArgProArgAspGluAspValAspLysProL 150
456 .....AATGAGCGCAAAAAGATCTCCAG 479
||||| :|||:|||||:|||||:|||||:|||||:
150 yalysIleLysThrLysAspThrLysLysGluLysArgLysLeuGlu 166
480 TGTACGGCCCTGTGAGATTAGACTTGCTCTGGAAGCCTTCACAG 529
||| :|||:|||||:|||||:|||||:|||||:
167 GluGluGluAspGlyLysLeuLysLysProLysAsnLysAspLys 183
530 CTCTTTAGCCAAACAAGAAAGATAGAACAGACCCCTTCAGAAAGCTT 579
||||| :|||:|||||:|||||:|||||:|||||:
183 sLysValProGluProAspAsnLysLysLysLysProLysGluGlu 200
580 TGAATCAA.....CTG 590
||||| :|||:|||||:|||||:|||||:|||||:
200 LglLysLysTrpLysTrpTrpGluGluArgLysTrpProGluLys 216
591 ATGAGCAATTCAGAGAAAGATCCAAAGTCTTCTTTCATTCCTGT 640
||| :|||:|||||:|||||:|||||:|||||:
217 TrpLysPheLeuGluHisLysGlyPro.....225
641 GACTGATTTTATGCTCTGCTACTCCATGATC.....ATTAAAC 681
||||| :|||:|||||:|||||:|||||:|||||:
226 .....ValPheAlaProProLysLysProLeuProGluAsnValLysP 240
682 ACCCAATGATTTTAGTACCATGAAA.....707
||| :|||:|||||:|||||:|||||:|||||:
240 heTyrTrpAspLysValMetLysLysLeuSerProLysAlaGluVal 256
708 .....GAAAAGATCAAGAACATGACTATGATCCATGAGAGA 745
||||| :|||:|||||:|||||:|||||:|||||:
257 AlaThrPhePheAlaLysMetLeuAspHisGlyLysThrThrLysGlu 273
746 ACTAAAGATTAATTCGAAACTAATGTACTAATGATGATTTACATA 795
||||| :|||:|||||:|||||:|||||:|||||:
273 ePheArgLysAsnPhe.....278
796 AACCAAGACCAATTTATTAAGCTGCAAGAGCTGTTCACATCAGGA 845
||| :|||:|||||:|||||:|||||:|||||:
279 .....PheLysAspTrpArgLys.....284
846 ATGAAATTTCTAGCCAGAGAAAGATTCAGACCTGAAAGACAGATAGA 895
||||| :|||:|||||:|||||:|||||:|||||:
285 ...GluMetThrAsnGluLys.....As 292
896 CTTCATGCTGACTTCAGAAACTCGAAAGAGAAAGATGGAACAGACA 945
||||| :|||:|||||:|||||:|||||:|||||:
292 nIleIleThrAsnLeuSerLysCys.....AspPheThrGlu 305
946 CTTACAGAGTGGGAGAGAGAGAGCTGCTGACAGAGAGAGAGAGC 995
||||| :|||:|||||:|||||:|||||:|||||:
305 eLysSerLysThrLys.....310
996 TCTGAGATGCCGAGACACAGCCTTCAGAGTCCAGAGAAAGATTA 1045
||||| :|||:|||||:|||||:|||||:|||||:
311 .....AlaGluThrGluAlaArgLysGluMetSerLysGluGlu 324
1046 AAAGAAAGCAAGATATGCTTGAAGATAGTTAAAGCAATTAAT... 1091
||| :|||:|||||:|||||:|||||:|||||:
324 sLeuLysIleLysGluGluAsnGluLysLeuLysGluTyrGlyPheC 341
1092 .....TTAGAGAGAGAGACAGACAGCTTGACCCGATC...GTGAAGGA 1133

```

```

341 ysllmetaspahnshtlslsgluarqlllealasnphelysllleuPro 357
1134 TCTGCA.....GGAAAGCTGAC 1150
358 ProglyleuphearglylarygllysnhshProlysmetglymetleu 374
1151 CAGGGGG.....CTGTGAACAGTCAGTCGCAATTGTG 1182
374 sarqatrglmetProgluaspllelleasn.CysSerlyasapla 390
1183 AAGAAGAAACACAGATGGAACACAGCTGTG.....GA 1217
391 lysvalProserProProProgllyhslslystrplysgluvalarghslas 407
1218 CTTTCCTCCATCCTGTGATCCATTTGTAGAGACCAGCTACTCCTGTG 1267
407 pasnllysalThrtrpleuVal..... 414
1268 GAGACTGGCAATGACCACTGGAGACTTCAGTCGGAGTGAATCTTGC 1317
415 ..SertrpThrGlusnlelnclly...Serllelystrylleleu 429
1318 AGGGCTCAAGAGATMAAGAACAAAGTCACTCCAGTGTAT 1362
430 AsnProSerSerArgllelysglyglulysasptprglInlystr 444
seq_name: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:PCT-US93-07261-11

```

```

seq_documentation_block:
? Sequence 11, Application PC/TUS9307261
? GENERAL INFORMATION:
? TITLE OF INVENTION: PfEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
? NUMBER OF SEQUENCES: 23
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: John H. C. Blasdale
? STREET: One Giralda Farms
? CITY: Madison
? STATE: New Jersey
? COUNTRY: USA
? ZIP: 07940-1000
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: Macintosh 6.0.5
? SOFTWARE: Microsoft Word 5.1a
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US93/07261
? FILING DATE: 19930805
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/927,531
? FILING DATE: 07-AUG-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Blasdale, John H. C.
? REGISTRATION NUMBER: 31,895
? REFERENCE/DOCKET NUMBER: DX0288K
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 201-822-7398
? TELEFAX: 201-822-7039
? INFORMATION FOR SEQ ID NO: 11:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1588 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? PCT-US93-07261-11

```

```

alignment_scores:
  Quality: 147.50
  Ratio: 0.467
  Percent Similarity: 47.094
  Length: 671
  Gaps: 32
  Percent Identity: 19.523

```

```

alignment_block:
US-09-687-230-1 x PCT-US93-07261-11 ..
Align seg 1/1 to: PCT-US93-07261-11 from: 1 to: 1588

```

```

309 GAAGACAAAAGATCATGACAAACACAGAC.....AGAAAGCGGA 352
29 GlnlyslsasnsaplaaglInlysalalysaspleutrhlyslgluse 45
353 AAGAAGAAAGAAAGAGAGAACAGATTCACAGGGGAGGAGAAAGGGGANA. 401
45 rglnAspSerSerSerSerSerSerleuSerleuSerleuValasnnglyua 62
402 .....AAGGAGAGAGATTAG 419
62 lAleuLyglulysgluasnlysgluThrleuLylyslglulencu 78
420 GAGATMAAAGAACGAGATGAGACCGGGTGAGATGAGCAGAAA 469
79 Asnglnlysglulysglulglulysasnlysllelysasnasnaspl 95
470 AGATCTCCAGTGTACGCCCTGTGAGATTAGAC.....TTGCCCTCTG 513
95 uAlAleuLyasnlysglyasnAsplysaspslyslsllvalProl 112
514 AGAAGCTCTCAACAAGCTCTTTAGCAACAAGAAAGATGAAACAGACA 563
112 yslsPro..... 114
564 CCCCTCAAGAGCTTTGATCAACTGATGAGACATTCGACGAGAAAGA 613
115 .....GluservalglulysAspleuLyglulmetgluleuLygl 128
614 TCCAAGTCTTCTTTCATTTCTCTGTGACTGATTTATTATTCCTCTGCT 663
128 ulys..... 129
664 ACTCCATGATCATTTAAACACCAATGATTTAGTACATGAAAGAAAG 713
130 ..GlupheillelysglnhshleuLyasptprglulglulysglulys 145
714 ATCAAGAAACATGCTATCACTCATTA.....GAAGAACTMAAGATTA 757
146 ArgArqAsntrpIleuArgSerleuArgArgAsplyleuArgglul 162
758 CTTCAACTAATGTGTACTAATGACATGATTTACATTAACCGAAGACA 807
162 eglulnleuglulysleuasnlaeglInleuGluserAlalleasnglul 179
808 TTTATTATAAGCTGCMAAGAG.....CTGTTCACATCA 842
179 euLysgluArgArgAlaserArgArgProMetValylsMetGlnArg 195
843 GGAATGAATAATTCTTAGCCAGGAAAGATTCAGAGCTTACAGACAGCAT 892
196 GlyMetlys.....AspGluValAspGluThrIleLylysltyas 209
893 AGACTTCAGGCTGACTTCAGAAACCTGAAAGCAGAAA..... 932
209 paspGluGlnAlaIglulysasnlylThrlyAspGluGluIlelysaapl 226
933 .....GATGGA.....ACAGACACTCACAGAGTGG..... 959
226 yslLyaspGlyTyrgluGluIleValIgluThrlyPhenylglyMetArg 242
959 ..... 959
243 GluasAlaIleuglygluleuAspGluTyrgluGluArgTyrgluLyly 259
960 .....GAGACGAGAGCTGCTGCGACAGAGAGAGAGAG 993
259 sarqTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 276

```

```

994 ACTCTGAGATGCGAGACACAGCCCTTCAGAGTCCC..... 1031
      :::::  :::::  :::::  :::::  :::::  :::::
276 IuylsleuGlulwthrGlyTyrGlyPheargGlyLysPheProthr 292
      :::::  :::::  :::::  :::::  :::::  :::::
1032 ..... AGCAAGCAAAATATAAAGAAAGCAAAAGATATGCTTGA 1069
      :::::  :::::  :::::  :::::  :::::  :::::
293 ArgIleLeuValIysArgLysArgAsnLysGluGlnLysLysLeuLysG 309
      :::::  :::::  :::::  :::::  :::::  :::::
1070 AGTTAGTCTTAAAGCAATATATTTAGAGAGACAGACAGACGCTTGACC 1119
      :::::  :::::  :::::  :::::  :::::  :::::
309 uAspLysGluLysLysLeuIleAlaIleGluGluProAspArgLysLys 326
      :::::  :::::  :::::  :::::  :::::  :::::
1120 GCATC...GTGAGGAATCTGAGGAACCTGACACAGCGCTTGTGAAAC 1166
      :::::  :::::  :::::  :::::  :::::  :::::
326 yStIleLysLeuLysAspSerAspLysValValAlaProValAsnLys 342
      :::::  :::::  :::::  :::::  :::::  :::::
1167 AGTCAGTGGCAATTT.....GAAAGAGAAAACACAGATGAAACAGAC 1210
      :::::  :::::  :::::  :::::  :::::  :::::
343 AsnLysSerPheProAspLysPheArgAlaProAspLysLysArgTh 359
      :::::  :::::  :::::  :::::  :::::  :::::
1211 G.....TTGGACTTCTCCATCCTGTGATCCCATTTGTAGAG 1248
      :::::  :::::  :::::  :::::  :::::  :::::
359 rMetPheTyrArgLeuSerGluLeuPheProIleValProArgLysAsp 376
      :::::  :::::  :::::  :::::  :::::  :::::
1249 AGCCAGGCTACTGCGCTGTGAGACTGGATGACAACTGGAAAGACTTCA 1298
      :::::  :::::  :::::  :::::  :::::  :::::
376 sn.....GluLeuAlaValLysGlyAspSerMetAsp 386
      :::::  :::::  :::::  :::::  :::::  :::::
1299 TCTGAGTGAAT.....ACTTGCAGGGGTTCAAAGA 1330
      :::::  :::::  :::::  :::::  :::::  :::::
387 SerLysValAsnGlyLysLysLeuLysSerThrPheAsnProPheLys 403
      :::::  :::::  :::::  :::::  :::::  :::::
1331 CGATTAAGCAACAAAGTCACTCCAGTGTATATTCATTTATGAGGGCC 1380
      :::::  :::::  :::::  :::::  :::::  :::::
403 g...ArgArgAsnLysLeuLysGluArgLysMetGlnGluLeuHisLys 419
      :::::  :::::  :::::  :::::  :::::  :::::
1381 ACAGTCTTATGACACCCATATGACTCCACATTTGCCAATATACGCAAG 1430
      :::::  :::::  :::::  :::::  :::::  :::::
419 heLysLysAsnTyrLysLysTyrGlnLysLeuGluArgGluLysArg 435
      :::::  :::::  :::::  :::::  :::::  :::::
1431 GATGAT..... 1436
      :::::  :::::  :::::  :::::  :::::  :::::
436 GluAsnProAspGlyGluProLeuAsnThrProGluIleHisValIle 452
      :::::  :::::  :::::  :::::  :::::  :::::
1437 ....TCTGATTTAATC..... 1448
      :::::  :::::  :::::  :::::  :::::  :::::
452 gProSerLysPheMetAspLysGlyGluAsnLysSerAlaGlyHisProp 469
      :::::  :::::  :::::  :::::  :::::  :::::
1449 .....TATTCACCTATGGG.....GAAAGACTGTGATCTT 1478
      :::::  :::::  :::::  :::::  :::::  :::::
469 heLysTyrGlnProThrLysGlyLeuLysGluTyrGlnGluSerHisVal 485
      :::::  :::::  :::::  :::::  :::::  :::::
1479 CCAAGTGAATTCAGCATC...CATGAGTTTGGCCACAGCTGCCAAGATTA 1525
      :::::  :::::  :::::  :::::  :::::  :::::
486 SerLysAspTyrGlnLeuGlnHisGluProProThrLysLeuProGlu 502
      :::::  :::::  :::::  :::::  :::::  :::::
1526 TCCG.....TATGTCATGCGCAGATAGTTTACTGGAT..... 1556
      :::::  :::::  :::::  :::::  :::::  :::::
502 rGluLysGluHisValSerArgGluTyrGlnLeuAspHisGluProPro 519
      :::::  :::::  :::::  :::::  :::::  :::::
1557 .....GTTTAAACAAAAGAGGCGCATTCAGGACCTACAA... 1592
      :::::  :::::  :::::  :::::  :::::  :::::
519 hLysLeuProGluTyrGluLysGlnHisValSerArgGluTyrGlnLeu 535
      :::::  :::::  :::::  :::::  :::::  :::::
1593 .....GAGATGAGATGTTCATTTGCGCTGAA...GATGAAGGCCATAGTAG 1633
      :::::  :::::  :::::  :::::  :::::  :::::
536 AspAsnGluValAlaArgAspGluLeuProGluTyrGluLysGlnHisVal 552
      :::::  :::::  :::::  :::::  :::::  :::::
1634 G.....ACACTTGAC.....ACAGGAAAGAAATGG 1659
      :::::  :::::  :::::  :::::  :::::  :::::
552 rArgGluTyrGlnLeuAspAsnGluLysProSerThrLeuLysGluTyr 569
      :::::  :::::  :::::  :::::  :::::  :::::
1660 AGCAGATTACAGAAATA.....GAGCCACCA 1685

```

```

      :::::  :::::  :::::  :::::  :::::  :::::
569 spGln...ThrGluLeuAlaLysGlyLysAspIleThrAsnLysProHis 584
      :::::  :::::  :::::  :::::  :::::  :::::
1686 GGGCGTTTGACTCCAGTCACTCAAGACAGGCTCATAGCCCTGAAGACAGT 1735
      :::::  :::::  :::::  :::::  :::::  :::::
585 GluSerValAspGluTyrAspGlnThrGluLeuAlaLysGlyLysAspI 601
      :::::  :::::  :::::  :::::  :::::  :::::
1736 AACAAATTTGGCGTCCAGTGAAGTTTGTGACTGTGAAAGACTGAAA 1785
      :::::  :::::  :::::  :::::  :::::  :::::
601 eThrAsn.....LysProHisGluSerValAspGluTyrAspGlnSer 616
      :::::  :::::  :::::  :::::  :::::  :::::
1786 TATTCAGAAAGAACTTGATGAGACACACAGATTTGCTCAGGGAACCTCCAG 1835
      :::::  :::::  :::::  :::::  :::::  :::::
616 IuLeuAlaLysGlyLysAspIleThrAsnLysProHisGluSerValAsp 632
      :::::  :::::  :::::  :::::  :::::  :::::
1836 GAA.....GCCACAGATGAACGTTTGAGACACACAGACC 1867
      :::::  :::::  :::::  :::::  :::::  :::::
633 GluTyrAspGlnThrGluLeuAlaLysGlyLysGluValThrAsnLysPr 649
      :::::  :::::  :::::  :::::  :::::  :::::
1868 CCTGGGAAACATG 1880
      :::::  :::::  :::::  :::::  :::::  :::::
649 OHISGluAsnLeu 653

```


168 AAGAACACAGACAGACAAAGTCGGACAAACCTCTACAGGAGTATGT 217
|||||
18 LysLysHISLysLysHISLysSerAspLysHISLeuTYrClnGlnLysVala 34
218 AAGAAAGCCCTGAAGCTGTCTCTCAAGTAGAGGAAAGCAAGTCAAGC 267
|||||
34 LcLysProLeuLysLeuValLcLysValLcLysLysValLcLysValLcHIS 51
268 AACTCTCAGCGGACAGTCGGGACAGCTCCAGCTCTTGAGACAA 317
51 LcLeuSerThrGlySerSerGlyHISAspSerSerLeuPheGlnAspLys 67
318 AACGATCATGACAAACACAGAGAGAGAAAGCGGAAAGAAAGAAAGAG 367
68 AsnAspHISAspLysHISLysAspArgLysArgLysLysArgLysLysG 84
368 AAGAAAGCAGATTCCAGGGGAAAGAAAGGAGAAACGAGAGAGATTA 417
84 YcLysLysGlnLcLeuProGlyGlnGlnLysGlnLysArgLysArgLysVal 101
418 AAGAGATATAAAGAGACGAGATCGAGACCGGTCGAGATAGCGCAGAA 467
101 YcGlnAspLysLysLysArgAspArgAspArgValGlnLysGlnLysValGln 117
468 AAGATCTCCAGTGTACGCCCTGTGAGATTAGACTTGCTCTCTGAGAA 517
118 LysAspLeuGlnLysHISAlaProValArgLeuAspLeuProGlnLys 134
518 GCCTCTCAGACGCTCTTAGCCAAACAGAGAGAGATAGAGAGACGCCC 567
134 sProLeuThrSerSerLeuAlaLysGlnGlnLysValGlnGlnLysThrPro 151
568 TTCAGAGAACCTTGAATCACTGATGAGACAAATTCGAGAGAAAGATCCA 617
151 euGlnGlnLcLeuAsnGlnLcLeuMetArgGlnLcGlnLysAspPro 167
618 AGTGGTTTCTTCATTTCTGTGACTGATTTTATTTGCTCCGCGCTACGC 667
168 SerAlaPhePheSerPheProValThrAspPheIleAlaProGlnLysIse 184
668 CATGATCATTTAAACACCCAAATGATTTTGTACCATGAGAAAGAAAGATA 717
184 rMetIleIleLysHISProMetLAspPheSerThrMetLysGlnLysIle 201
718 AGAACAATGACTATCATGCTCATAGAGAACTAAAGATTAACCTCAAACTA 767
201 YSAsnAsnAspTyrGlnSerIleGlnGlnLcLeuLysAspAsnPhelysLeu 217
768 ATGTGTACTAATGCCATGATTTTACAAATAACAGAGACCATTTATTATA 817
218 MetCysThrAsnAlaMetIleTyrAsnLysProGlnThrIleTyrTyrIy 234
818 AGCTGCAAAAGAGCTGTGCACTCAGAGATGAATAATTCTTACCGAGGAA 867
234 sAlaAlaLysLysLysLeuLeuHISerGlyMetLysIleLeuSerGlnGln 251
868 GAATTCAGAGCCTGAAGCAGACATAGACTTCATGGCTGACTTGAGAAA 917
251 rGlnLcGlnSerLeuLysGlnSerIleAspPheMetAlaAspLeuGlnLys 267
918 ACTGCAAGCAGAAAGATGGAACAGACACCTCAGACAGAGTGGGAGAGCG 967
268 ThrArgLysGlnLysAspGlyThrAspThrSerGlnSerGlnLysAspG 284
968 AGGCTGCTGGCAGAGAGAGAGAGACTCTGAGATGCGGAAAGCAGACAG 1017
284 YcLysYsrTrpGlnArgGlnArgGlnAspSerGlyAspAlaGlnAlaHIS 301
1018 CCTTCAAGAGTCCAGCAAAAGAAATTAAGAAAGAGCAAAAGATATGCTT 1067
301 lAphelySerProSerLysGlnAsnLysLysLysAspLysAspMetLeu 317

1068 GAAGATTAAGTTTAAAGCAATAATTTAGAGAGAGAGAGAGAGACCTTGA 1117
|||||
318 GlnAspLysPheLysSerAsnAsnLeuGlnArgGlnGlnGlnLcLeuAs 334
1118 CCGCATCGTGAAGAAATCTCGAGAGAAAGCTGACCGCGGCTTGTAACA 1167
334 rArgIleValLysLysGlnSerGlyLysLysLeuThrArgArgLeuValAsn 351
1168 GTCAGTCCGCAATTGAAAGAGAAAGAAACAGATGGAACAACAGACGTTGG 1217
351 erGlnLysGlnPheGlnArgArgLysProAspGlyThrThrThrLeuGly 367
1218 CTCTCCATCTCTGTGAATCCATTTAGAGAGACCGACGCTACTGCTGGT 1267
368 LeuLeuHISProValAspProIleValGlnGlnProGlyTyrCysProVa 384
1268 GAGACGTGGGAATGACAACCTGAGAGACTTCAGCTGGAGGCAATCTTGC 1317
384 LArgLeuGlyMetThrThrGlyArgLeuGlnSerGlyValAsnThrLeuG 401
1318 AGCGTTCAAGAGATATAAAGAGACAAAGTCACTCCAGTATATATTGG 1367
401 lncGlyPheLysGlnAspLysArgAsnLysValThrProValLeuLysLeu 417
1368 AATTATGGCCCTTACAGTTCTTATGCAACCGCATTAATGACTCCACATTCG 1417
418 AsnTyrGlyProTyrSerSerTyrAlaProHISLysAspSerThrPheAl 434
1418 AATATCAGACAGATGATTCGATTTAATCTATTCACCTAGGGGAAG 1467
434 AsnIleSerLysAspAspSerAspLeuIleTyrSerThrTyrGlnGln 451
1468 ACTCTGATCTTCCAACTGATTTTCAGCATCCATCAGTTTGGCCAGCTGC 1517
451 sPserAspLeuProSerAspPheSerIleHISGlnPheLeuAlaThrCys 467
1518 CAAGATTATCCGATATGTCATGGCAGATAGTTTACTGAGATGTTTAAACA 1567
468 GlnAspTyrProTyrValMetAlaAspSerLeuLeuAspValLeuThrLys 484
1568 AGAGGGCAGATTCAGAGACCTTCAAGAGATGGAGATGTCATGGCCGGAAG 1617
484 sGlyGlyHISerArgThrLeuGlnLcLeuMetGlnMetSerLeuProGln 501
1618 ATGAAAGCCACTACTAGACACTTGACACAGAGAAAGAAATGAGCAGATT 1667
501 sProGlnLysHISThrArgThrLeuAspThrAlaLysLcMetGlnGlnIle 517
1668 ACAGAAATGAGACCCACAGGCGTTTGGACTCCAGTACTCAAGACAGGCT 1717
518 ThrGlnValGlnProProGlyArgLeuAspSerSerThrGlnAspArgLe 534
1718 CATAGCGCTGAAGAGCAGTAACAATAATTTGGCGTCCAGTTGAAGTTTGG 1767
534 uIleAlaLeuLysAlaValAlaThrAsnPhelyValProValGlnValPhe 551
1768 ACTCTGAAGAACCTGAATATTTCCAGAGAAACCTTGATGAGACCAACAG 1817
551 sPserGlnGlnLysAlaGlnLcLeuPheGlnLysLysLeuAspGlnThrThrArg 567
1818 TTGCTCAGGGAATCCAGAGACCCCAAGATGAACGTTTGAACACAGACC 1867
568 LeuLeuArgGlnLeuGlnGlnLysAlaGlnAsnGlnLysArgLeuSerThrArgPr 584
1868 CCTGGGACATGATGCTGTCTCTGGGTCCTTCATC .AGAGAAATGCATC 1916
584 oProProAsnMetIleCysLeuLeuGlyProSerTyrArgGlnMetHISL 601
1917 TTGCTGAACAAGTGAACCAATTAATCTTAAAGAAAT .GCACAGCAAGTAACT 1965
601 euAlaGlnGlnValThrAsnAsnLeuLysGlnLeuAlaGlnGlnValThr 617
1966 CCAGGTGATATCGTAAGCACGTATGGAGTTGCAAAAGCAATGGGATTTTC 2015

|||||
618 ProGlyAspIleValSerThrItyrGlyValArgLysAlaMetGlyIleSe 634
2016 CATTCCCTTCCCCCTCATGAGAAACACTTTGTGATTGACAGAGACA 2065
|||||
634 rIleProSerProValMetGluAsnAsnPheValAspLeuThrGluAspT 651
2066 CTGAAAGAACTTAAAGACGATGTTGATGACTGTCGACCTGGGAGAGT 2115
|||||
651 htrGluGluProLysLysThrAspValAlaGluCySGlyProGlyGlySer 667
seq_name: /SIS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAU16626
seq_documentation_block:
ID AAU16626 standard; Protein; 718 AA.
XX
XX AAU16626;
DT 07-NOV-2001 (first entry)
XX
DE Human novel secreted protein. Seq ID 1579.
XX
XX Human; immunosuppressive; antiarthritic; antirheumatic;
KM cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KM neuroprotective; antibacterial; virucide; fungicide; opthalmological;
KM vulnery; secreted protein; rheumatoid arthritis;
KM hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KM cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KM nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KM corneal infection; wound healing; epithelial cell proliferation;
KM skin ageing; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN WC20015322-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01341.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.

PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.


```

1018 CCTTCAGAGTCCACGAAAGAAATTAAGAAAGACAAAGATATGCTT 1067
|||||
333 IAPhElySerProSerLysGluAnLysLysAspLysAspMetLeu 369
|||||
1068 GAAGATTAAGTTAAAGCAATATATTAGAGAGACGAGACACTTGA 1117
|||||
370 GluAspLysPheLysSerAsnAsnLeuGluArgLugLugLugLeuAs 386
|||||
1118 CCGCATCTGGAAGATCTGGAGAAAGCTGACAGCGCGCTTGTAACA 1167
|||||
386 pArgLLeuValLysGluSerGlyLysLeuThrArgArgLeuValAsn 403
|||||
1168 GTCACTGGGAATTTGAAGAAAGAAACAGATGGAACAACGCTTGGA 1217
|||||
403 eGlnCysGluPheGluArgLysProAspGlyThrThrThrLeuGly 419
|||||
1218 CTTCCTCATCCCTGTGATCCCATTTAGAGAGACGAGGCTACTGCTGT 1267
|||||
420 LeuLeuHisProValAspProLeuValGlyLugProGlyTyrCysProVa 436
|||||
1268 GAGACTGGGAATGACACACTGGAAAGACTTCACTGGAGTGAATACTTGC 1317
|||||
436 IArgLeuGlyMetThrThrGlyArgLeuGlnSerGlyValAsnThrLeuG 453
|||||
1318 AGGGGTTCGAAGAGATTAAGGAACAAAGTCACCTCCAGTGTATATTTC 1367
|||||
453 IngLyPheLysGluAspLysArgAsnLysValThrProValLeuTyrLeu 469
|||||
1368 AATTATGGCCCTACAGTTCATTATGACACCGCATTTATGACTCCACTTTC 1417
|||||
470 AsnTyrGlyProTyrSerSerTyrAlaProHisTyrAspSerThrPheAl 486
|||||
1418 AATATTCGCAAGATGATTTGATTATATCTATTCAAACCTATGGGAG 1467
|||||
486 AsnLLeuSerLysAspAspSerAspLeuLLeuTyrSerThrTyrGlyGlu 503
|||||
1468 ACTGATCTCTCCAGATGATTTCAGACATCCATGATTTTGGCCACGTC 1517
|||||
503 spSerAspLeuProSerAspSerIleHisGluPheLeuAlaThrCys 519
|||||
1518 CAAGATTATCCGTATGTCATGGCAGATAGTTTACTGATGTTTAAACAA 1567
|||||
520 GlnAspTyrProTyrValMetAlaAspSerLeuLeuAspValLeuThrLy 536
|||||
1568 AGGAGGGATTCAGAGACCTTACAAGAGATGAGATGTCATGCTGTGAG 1617
|||||
536 sGlyGlyHisSerArgThrLeuGlnGluMetGluMetSerLeuProGlu 553
|||||
1618 ATGAAGGCGCACTAGACACTTGACACAGAAAGAAATGAGACAGATT 1667
|||||
553 spGluGlyHisThrArgThrLeuAspThrAlaLysGluMetGlu...Ile 568
|||||
1668 ACAGAAAGTAGACCCACGAGGCGTTTGGACTCCAGTACTCAAGACAGCT 1717
|||||
569 ThrGluValGluProProGlyArgLeuAspSerSerThrGlnAspArgLe 585
|||||
1718 CATGACCGCTGAAGACGATACAAATTTGGCGTCCAGTTGAAGTTTGG 1767
|||||
585 uIleAlaLeuLysAlaValThrAsnPheGlyValProValGluValPhe 602
|||||
1768 ACTCTGAAGAGCTGAATATTTCCAGAAAGAACTTGATGAGACACAGCA 1817
|||||
602 spSerGluGluAlaGluIlePheGlnLysLysLeuAspGluThrThrArg 618
|||||
1818 TTGCTCAGGGAACCTCAGGAAGCCACCAATGAACTTTGAGACACAGACC 1867
|||||
619 LeuLeuArgGluLeuGlnGluAlaGlnAsnGluArgLeuSerThrArgP 635
|||||
1868 CCCGGGAACATGATCTCTCTCTGGGCTCCTCATC AGAAGATGCATC 1916
|||||
635 oProProAsnMetIleCysLeuLeuGlyProSerTyrArgGluMetHisL 652

```

```

1917 TTGCTGAACAAGTGACCAATATCTTAAAGATT GCACAGCAAGTAAC 1965
|||||
652 euAlaGluGlnValThrAsnAsnLeuLysGluLeuAlaGlnGlnValThr 668
|||||
1966 CCAAGTATATCGTACGACGTATGAGTTCGAAAGCAATGGGATTC 2015
|||||
669 ProGlyAspIleValSerThrTyrGlyValArgLysAlaMetGlyIleSe 685
|||||
2016 CATTCCTTCCCCGTCATGAGAAACAACTTGTGGATTGACAGAAACA 2065
|||||
685 rIleProSerProValMetGluAsnAsnPheValAspLeuThrGluAsp 702
|||||
2066 CTGAAGAACTTAAAAAGACGATGTTGATGAGTGTGACCTGTGTGAAGT 2115
|||||
702 hrgLugLugProLysLysThrAspValAlaGluCysGlyProGlyGlySer 718
|||||
seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.AAM37947
seq_documentation_block:
ID AAM37947 standard; Protein; 589 AA.
XX
AC AAM37947;
XX
DF 14-SEP-1998 (first entry)
XX
DE Phosphatidylinositol-3' kinase associated protein.
XX
KM Phosphatidylinositol-3' kinase associated protein; PI3K; PI3AP;
KW human; signal transduction; cell growth; cancer; restenosis;
XX therapy; diagnosis.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 151..313
FT /note="Bromodomain"
FT Region 516..589
FT /note="PI3K p85 binding region"
XX
PN W09820126-A1.
XX
PD 14-MAY-1998.
XX
PF 01-OCT-1997; 97WO-US15845.
XX
PR 01-NOV-1996; 96US-0030103.
XX
PA (ONXX-) ONXX PHARM INC.
XX
PI Braselmann S;
XX
DR WPI: 1998-286942/25.
XX
DR N-PSDB: AAV29267.
XX
PT New isolated phosphatidylinositol-3'kinase associated protein -
PT used to develop products for diagnosis and treatment of cell growth
PT disorders such as restenosis or cancer
XX
PS Claim 10; Page 40-41; 52pp; English.
XX
CC This polypeptide comprises human phosphatidylinositol-3' kinase
CC (PI3K) associated protein (PI3AP), a protein that binds to the
CC intermediate SH2 domain on the p85 regulatory subunit of PI3K, and
CC which exhibits a bromodomain. Its amino acid sequence was deduced
CC from a cDNA clone (see AAV29267) obtained from an HeLa library using
CC a yeast two-hybrid assay with PI3K p85 as bait. The invention
CC provides vectors containing nucleic acid sequences that encode
CC PI3AP or its fragments, host cells, methods for the expression of
CC PI3AP, and methods for using the products for the diagnosis and
CC treatment of cell growth disorders such as restenosis or cancer.
CC Also described is an assay for identifying agonists and antagonists
CC of PI3K regulation. These include mutant PI3APs that compete with
CC native PI3APs for binding to PI3K, antibodies, and nucleotide

```

CC sequences that can be used to inhibit or enhance PIPAK gene
 CC expression. Transgenic and knock-out animals are also described.
 XX
 50 Sequence 589 AA;

alignment_scores:
 Quality: 3073.00 Length: 589
 Ratio: 5.217 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-687-230-1 x AAM37947 ..

Align seg 1/1 to: AAM37947 from: 1 to: 589

```

162 ATGGGCAAGACACACAAAGCAAGCAATCGACAAACACCTTCAGAGCA 211
|||||
1 MetGlyLysLysHisLysLysHisLysSerAspLysHisLeuTyrGlu 17
212 GATGTGAGAGAGCCCTTGAGCTGGCTCCAAAGTAGAGGAGGAGCAAG 261
|||||
17 TyrValGluLysProLeuLysLeuValLeuLysValGlyLysGlu 34
262 TCACCGAAGCTTCACAGGAGCTCGGGCAGACTCCAGCTCTTCGAA 311
|||||
34 AlrHcLysLeuSerThrGlySerSerGlyHisAspSerLeuPheGlu 50
312 GACAAAAGATCATGACAAACACAGACAGAAACCGGAAAAAGGAAA 361
|||||
51 AspLysAsnAspHisAspLysHisLysAspArgLysArgLysArg 67
362 GAAGAGAGAGAGCAGATTCCAGGGGAGAGAAAAAGGGAGAAAAAGG 411
|||||
67 slySGlyGlyLysGlnIleProGlyGluGlyLysGlyArgLysArg 84
412 GACTTAAGAGGATTAAGAAAGACGAGATCGAGACCGGGTGGAGATG 461
|||||
84 rValLysGlyLysLysLysArgAspArgValGluAsnGlu 100
462 GCAGAAAAGATCTCAGTGTCCAGGCCCGCTGTGAGATTAGACTTGC 511
|||||
101 AlaGluLysProLeuGlnCysHisAlaProValArgLeuAspLeuPro 117
512 TGAGAGCCTCTCACAGCTCTTTAGCCAAACAAAGAGAGAGAGAGAG 561
|||||
117 OGlyLysProLeuThrSerSerLeuAlaLysGlnGluValGluGln 134
562 CACCCCTTCAGAGAGCTTGAATCACTGATGAGACATTCGACAGAAA 611
|||||
134 hrProLeuGlnGluAlaLeuAsnGlnLeuMetArgGlnLeuGlnArg 150
612 GATCAAGTCTCTTCTTCAATTCCTGTGACTGATTTTATTCCTCTG 661
|||||
151 AspProSerLalpheSerPheProValThrAspPheIleAlaProG 167
662 CTACTCCAGATCATTAACACCCCAATGCAATTTAGTACCATGAAGAA 711
|||||
167 yTyrSerMetIleIleLysHisProMetAspPheSerThrMetLys 184
712 AGATCAAGACATGACTATCAGTCAATAGAGACATTAAGAGATTC 761
|||||
184 ysiLeuAsnAsnAspTyrGlnSerIleGlnGluLeuLysAspAsp 200
762 AAATAATGTGTAATGCAATGATTTACATTAACACAGAGACATTTA 811
|||||
201 LysLeuMetCysThrAsnAlaMetIleTyrAsnLysProGluThrIle 217
812 TTATTAAGCTGCAAGAGCTGTTGACATCAGAGATGAAAAATTTTGA 861
|||||
217 rTyrLysAlaLysLysLeuHisSerGlyMetLysIleLeuSerG 234
862 AGAAAAGATTCAGAGCTGAGACAGACATAGACTTCATGGCTGACT 911

```

```

|||||
234 InGluArgIleGlnSerLeuLysGlnSerIleAspPheMetAlaAspLeu 250
912 CAGAAAAGCTGCAAGACAGAAAGATGGAACAGACACTCCACAGAGGG 961
|||||
251 GlnLysThrArgLysGlnLysAspLysThrAspThrSerGlnSerGly 267
962 GCAGGAGGCTCTGGCAGAGAGAGAGAGAGAGAGCTGTGAGATGCC 1011
|||||
267 uAspGlyGlyCysTrpGlnArgGluArgGluAspSerGlyAspAlaGlu 284
1012 CACAGCCTTCAGAGTCCACCAAGAAATTAAGAAAGAAACAAAGAT 1061
|||||
284 lAlaHisLalpheLysSerProSerLysGluAsnLysLysLysAsp 300
1062 ATGCTTGAGATTAAGTTTAAAGCAATATTTAGAGAGAGACAGAGCA 1111
|||||
301 MetLeuGluAspLysPheLysSerAsnLeuGluArgGluGlnGlu 317
1112 GCTTGACCCATCGTGAAGAAATCTGAGAGAAAGCTGACAGCGGCT 1161
|||||
317 LeuAspArgIleValLysGluSerGlyLysLeuThrArgArgLeu 334
1162 TGACAGCTGAGTCCGAATTTGAAGAAAGAAACCGATGGAACAGAG 1211
|||||
334 AlaAsnSerGlnCysGluPheGluArgArgLysProAspLysThrThr 350
1212 TTGGCACTTCCATCCTGTGATCCCATTTGAGAGAGCCAGGCTACT 1261
|||||
351 LeuGlyLeuLeuHisProValAspProIleValGlyLysProGlyTyr 367
1262 CCGTGAGACTGGGAGATGACAACTGGAAGACTTCAGTGTGAGTGA 1311
|||||
367 sLeuValArgLeuGlyMetThrThrGlyArgLeuGlnSerGlyValAsp 384
1312 CTTTGAGGGGTTCAAGAGAGATTAAGAAAGAAAGTCACTCAGTGA 1361
|||||
384 hrLeuGlnLysPheLysGluAspLysArgAsnLysValThrProVal 400
1362 TATTTGAATTAATGGCCCTPACAGTTCTTATGACCGCATTAATG 1411
|||||
401 TyrLeuAsnTyrGlyProTyrSerSerTyrAlaProHisLysAsp 417
1412 ATTTGCAATATTCAGAGAGATTCATGATTAATCTTCAACCTAG 1461
|||||
417 rPheAlaAsnIleSerLysAspAspSerAspLeuIleTyrSerThr 434
1462 GGGAGACTCTGATCTTCCAGATGATTTGAGATTCATGATGATTT 1511
|||||
434 LysLysAspSerAspLeuProSerAspPheSerIleHisGluPhe 450
1512 AGTGGCAAGATTAATCCGTAATGTCATGCGCAGATAGTTTACTG 1561
|||||
451 ThrCysGlnAspTyrProTyrValMetAlaAspSerLeuLeuAsp 467
1562 AACAAAAGGAGGCAATTCAGAGCCCTACAAAGATGAGATGCTCAT 1611
|||||
467 uThrLysGlyGlyHisSerArgThrLeuGlnGlnMetGluMetSer 484
1612 CTGAGATGAAGGCCATACTAGACACTTGACACAGAAAAAGAAATG 1661
|||||
484 rGluLysArgLysGlnHisThrArgThrLeuAspThrGlyLysGlu 500
1662 CAGATTACAGAAATAGACACAGAGGCGTTTGAGACTCCAGTACTGA 1711
|||||
501 GlnIleThrGluValGluProGlyArgLeuAspSerSerThrGln 517
1712 CAGGTCATAGCGCTGAAAGCAGTAACAAATTTGGGCTCCAGTTGA 1761
|||||
517 ArgLeuIleAlaLeuLysAlaValThrAsnPheGlyValProVal 534
1762 TTTTGACTCTGAAGAGCTGAATATTCAGAGAAAGAACTTGATGAG 1811
|||||

```

534 a1PheAspSerGluGluAlaGluIlePheGlnLysLysLeuAspGluThr 550
 1812 ACCAGATTCTCAGGAGACTCCAGAGAGCCAGAAATGACGTTTGAGCAC 1861
 551 ThrArgLeuLeuArgGluLeuGlnGluAlaGlnAsnGluArgLeuSerTh 567
 1862 CAGACCCCTGGGACATGATCTGTCTTGGGTCCCTCATCAGAGAAAT 1911
 567 rArgProGlyAsnMetIleCysLeuLeuGlyProSerSerGluLysC 584
 1912 GCATCTTGGCGACACAG 1928
 584 yb1LeuLeuAsnLys 589

seq_name: /SID51/gcdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AA93712

seq_documentation_block:
 ID AA93712 standard: Protein; 405 AA.

AC AA93712;
 DT 06-NOV-2001 (first entry)
 DE Human polypeptide, SEQ ID NO: 3652.
 KM Human; full length cDNA; cDNA synthesis; oligo-capping.
 OS Homo sapiens.
 PN EP1130094-A2.
 PD 05-SEP-2001.
 PF 07-JUL-2000; 2000EP-0114089.
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 PA (HELI-) HELIX RES INST.
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 DR WPI: 2001-524255/58.
 DR N-PSDB: AA934661.
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 PS Claim 8; SEQ ID NO 3652; 1380bp + sequence listing; English.
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a polypeptide
 CC encoded by a full length human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 SQ Sequence 405 AA;

alignment_scores:
 Quality: 2013.50 Length: 406
 Ratio: 5.034 Gaps: 1
 Percent Similarity: 98.522 Percent Identity: 98.276

alignment_block:
 US-09-687-230-1 x AA93712 ..
 Align seq 1/1 to: AA93712 from: 1 to: 405

900 ATGGCTGACCTTGCAGAAACCTCCGAAAGCAAGAAATGTAACAGACCTC 949
 1 MetAlaSprLeuGlnLysThrArgLysGlnLysAspLysThrAspThrSe 17
 950 ACAGATGGGAG 999
 17 rGlnSerGlyGluAspGlyCysTrpGlnArgGluArgGluAspSerG 34
 1000 GAGATGCCGAAAGCAGACAGCCTTCAAGAGTCCAGCAAAAGAAATTAAGA 1049
 34 LysArgAlaGluAlaHisAlaPheLysSerProSerLysGluAsnLys 50
 1050 AAAGCAAAAGATATGCTGAAGATTAAGTTAAAGCAATATTAAGAGAG 1099
 51 LysAspLysAspMetLeuGluAspLysPheLysSerAsnAsnLeuGluAr 67
 1100 AGACAG 1149
 67 gGluGlnGlnGlnLeuAspArgIleValLysGlnSerGlyLysLeu 84
 1150 CCAGCGGCTTGTGACAGATGCGAATTTGAAAGAAAGAAACAGAT 1199
 84 hrArgArgLeuValAsnSerGlnCysGluPheGluArgGlyProAsp 100
 1200 GGAACAG 1249
 101 GlyThrThrThrLeuGlyLeuLeuHisProValAspProIleValGly 117
 1250 GCCAGGCTACTGCTGTGAGACTGGGAATGACAAGTGAAGAGACTTCACT 1299
 117 uProGlyLysCysProValArgLeuGlyMetThrThrGlyArgLeuGln 134
 1300 CTGAGATGAATCTTTGAGAGGGGTTCAGAGAGATTAAGAGAAAGAAATC 1349
 134 erGlyValAsnThrLeuGlnGlyPheLysGluAspLysArgAsnLysVal 150
 1350 ACTCCAGTGTATATTGAAATTAAGGGCCCTACAGTTTATGACCCACA 1399
 151 ThrProValLeuTyrLeuAsnTyrGlyProTyrSerSerTyrAlaProH 167
 1400 TTATGACTCCACATTTGCAAAATATCAGCAAGAGATTTCTGATTAATCT 1449
 167 styAspSerThrPheAlaAsnIleSerLysAspAspSerAspLeuIleT 184
 1450 ATTCAACCTATGGGAGAGACTGTATCTTCCAGAGATTTAGCATCAT 1499
 184 ySerThrTyrGlyGluAspSerAspLeuProSerAspPheSerIleHis 200
 1500 GAGTTTGGCCAGAGTCCAGATATATCCGATGATGAGGCGAGATGATT 1549
 201 GluPheLeuAlaThrCysGlnAspTyrProTyrValMetAlaAspSerLe 217
 1550 ACTGGATGTTTAAACAAAGAGAGGCAATTCAGAGCCTTACAGAGATGG 1599
 217 uLeuAspValLeuThrLysGlyLysHisSerArgThrLeuGlnGluMetG 234
 1600 AGATGTATTTGCTGAAGATGAAGCCATTAAGAGCACTTGACACAGAGA 1649
 234 lMetSerLeuProGluAspGluGlnHisThrArgThrLeuAspThrAla 250
 1650 AAAGAAATGAGAGAGATTAAGAAATAGAGCCAGAGGCGTTTGAGACTC 1699
 251 LysGluMetGlu...IleThrGluValGluProProGlyArgLeuAspSe 266
 1700 CAGTACTGACAGAGAGCTCATAGCGCTGAAGACAGTAAACAAATTTGGCG 1749
 266 rSerThrGlnAspArgLeuIleAlaLeuLysAlaValThrAsnPheGlyY 283

```

1750 TTCCAGTTGAGTCTTTTGGCTGAGAGAGCTGAATATATCCAGAGAAA 1799
|||||
293 aPrivaGIuValIhPheaspSerGIuValIhLeGInIlePheGInIlyS 299
|||||
1800 CTGTGAGAGACCAACGATTGCTCAGAGAACTCCAGAGAGCCAGAAATGA 1849
|||||
300 LeuAaPGLuThrThrArgLeuLeuArgGIuLeuGInIuValIhAsn 316
|||||
1850 ACGTTTGAGACCAACGAGCCCGCTGGGAGATGATGTCCTTGGGTCCT 1899
|||||
316 uArgLeuSerThrArgProProProAsnMetIleCysLeuLeuGIyProS 333
|||||
1900 CATC AGAGAAATGCATCTGTGTGAACAGTGAACCAATATCTTAAAGAA 1948
|||||
333 eTtYrArgIuMetHISLeuAlaGIuGInIuValThraSnaSneulysGIu 349
|||||
1949 TTGCACAGCAAGTAATCTCCAGTGTATATCTAAGCAGCTATGAGATGG 1997
|||||
350 LeuAlaGIuGInIuValThrProGIyAspIleValSerThrTyGIyAla 366
|||||
1998 AAAGCAATGGGATTTCCATTCCCTCCCGCTCATGAGAAAACACTTGG 2047
|||||
366 gIySaIaMetGIyIleSerIleProSerProValMetGIuAsnAsnPh 383
|||||
2048 TGGATTGACAGAGACACTGAGAAAGCACTAAAGAGAGATGTGATGG 2097
|||||
383 aIAspLeuThrGIuAspThrGIuGIuProIySylsThrAspAlaGIu 399
|||||
2098 TGTGACCTGTGTGAAGT 2115
|||
400 CysGIyProGIyGIySer 405
|||||
seq_name: /SIDS1/9c9data/hold-geneseq/gene-seq-emb1/AA2001.DAT: AAB93765
seq_documentation_block:
ID AAB93765 standard; Protein: 351 AA.
XX
AC AAB93765;
XX
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:13461.
XX
KM Human; primer: detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 13461; 2537bp + CD ROM; English.
XX

```

```

CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 351 AA:

```

```

alignment_scores:
Quality: 1716.50      Length: 352
Ratio: 4.975          Gaps: 1
Percent Similarity: 98.011 Percent Identity: 97.727

```

```

alignment_block:
US-09-687-230-1 x AAB93765 ..

```

```

Align seg 1/1 to: AAB93765 from: 1 to: 351

```

```

1062 ATGCTTGAAGATTAAGTTTAAAGCAATATTTAGAGAGAGAGAGAGCA 1111
|||||
1 MetLeuGIuAspIySylsPheIySerAsnAsnLeuGIuArgGIuInIuGI 17
1112 GCTTGACCGCATGCTGAAGAAATCTGAGAGAAAGCTGACAGGCGGCTGG 1161
|||||
17 nIeuAspArgIleValIyGIuGIuSerIyGIySleuThrArgArgLeuV 34
1162 TGAACAGTCAGTCCGAATTTGAAAGAAAGAAACAGATGGAACAGACAG 1211
|||||
34 aIAsnSerGIuGIuGIuPheGIuArgArgIyProAspGIyThrThr 50
1212 TTGGGACTTCTCCATGCTGTGATCCCATTTGAGAGAGAGAGGCTAGTG 1261
|||||
51 LeuGIyLeuIleuHISProValAspProIleValGIyGIuProGIyTyty 67
1262 CCTGTGAGAGACTGGCAATGACAACCTGAGAGACTTCAGTGTGAGTGAATA 1311
|||||
67 sProValArgLeuGIyMetThrThrGIyArgLeuGIuInsSerGIyAlaAsn 84
1312 CTTTGACGGGGTTCAAGAGATTAAGAGAAACAGCAACTCCAGTGTGA 1361
|||||
84 hIreuGIuInIyPheIySylsGIuAspIyArgAsnIyValThrProValIe 100
1362 TATTGGAATTATGGGCGCCACAGTCTTATGACACCGCATTAATGACCCAC 1411
|||||
101 TyIreuAsnTyGIyProIySerSerTyIAlaProHISTyIAspSerThr 117
1412 ATTTCGAATATACCAAGAGATGATGATGATTAATCTAATCAACTATG 1461
|||||
117 rHelaAsnIleSerIyAspAspSerIleuIeTyIerSerThrTyGIc 134
1462 GGGAAAGCTGTGATCTTCCAGTGAATTTTCAGCATTCAGATGATTTGGCC 1511
|||||
134 TyGIuAspSerAspLeuProSerAspPheSerIleHISGIuPheLeuAla 150
1512 ACGTGCAAGATTATCCGTAATGTCAATGAGATAGTTCAGATGATTTT 1561

```

```

151 TThcysGlnhspTryProtyValMetAlaAspSerLeuLeuaspValle 167
1562 AACAAAGAGGAGGATTCAGAGACCCCTACAGAGATGAGATGTCATTC 1611
167 uThrlYsGlyGlyHIsSerIrgThrlengInglumetGImetSerleup 184
1612 CTGAGAGATGAGAGCCATCTAGAGACCTTGACACAGAGAAAGAAAGGAG 1661
184 roclYaspGluGlyHIsThrArgThrlEusprhAlaLysGImetGlu 200
1662 CAGATTCAGAGATGAGAGCCAGAGGCGCTTGAGCTCCAGTCTCAGA 1711
201 ...lIerhGluValGluProGlyArgLeuAspSerSerThrgInas 216
1712 CAGGCTCATAGCGCTGAGAGACAGTAAATTTGGCGTTCAGATTGAG 1761
216 pArgleuIleAlaLeuYsAlaValThrasnpheglyValProValGlu 233
1762 TTTTGTACTGTGAAGAGCTGAATATTCAGAGAAATCTGATGAGACC 1811
233 alpheAspSerGluGluValaGluIlephGlnLysLysLeuAspGluThr 249
1812 ACCAGATTCGTCAAGGAGCTCCAGAGAGCCAGAGATGAGATGAGTTC 1861
250 ThraArgleuLeuArgGluLeuGlnGluValaGlnsnGlnuArgleuSerTh 266
1862 CAGACCCCTGGGAGACATGATCTGTCTCTGGGCTCCCTCATC. AGAGAA 1910
266 rArgProProProAsnMetIleCysLeuLeuGlyProSerTyrArgGlu 283
1911 TGCATCTGCTGAGACAGTACCAATATCTTAAAGATT. GCACAGCAA 1959
283 ethIsleuAlaGluGluGlnValThrasnAsnLeuYsGluLeuAlaGln 299
1960 GTAACTCCAGTATGATCTGAGAGATGAGATGAGTGAAGAGATGGG 2009
300 ValThrProGlyAspIleValSerThTyrGlyAlaArgYsAlaMetG 316
2010 GATTTCCATTCCTTCCCGCTCATGAGAAACACTTTGGATTTGACAG 2059
316 yIleSerIleProSerProValMetGluAsnAsnPhelValAspLeuThrg 333
2060 AAGACACTGAGAGACTTAAAGAGAGATGTTGATGAGTGCACCTGGT 2109
333 lAspThrgIuGluProLysLysThraspValAlaGluCysGlyProGly 349
2110 GGAAGT 2115
350 GlySer 351
seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AA040621
seq_documentation_block:
ID AA040621 standard: Protein: 605 AA.
XX
XX AA040621:
AC
XX
XX 22-OCT-2001 (first entry)
DX
XX
XX Human polypeptide SEQ ID NO 5552.
DE
XX
XX Human: nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
XX Homo sapiens.
OS
XX
XX WO200153312-A1.
XX

```

```

PD 26-JUL-2001.
XX
XX 26-DEC-2000; 2000MO-US34263.
PF
XX
XX 21-JAN-2000; 2000US-0488725.
PR
XX 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
DR
XX N-PSDB; AA159777.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 5552; 10078bp; English.
PS
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AA038642-AA042213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 605 AA:
SQ
XX
XX
XX alignment_scores:
XX Quality: 927.50 Length: 627
XX Ratio: 2.268 Gaps: 18
XX Percent Similarity: 65.231 Percent Identity: 34.769
XX
XX alignment_block:
XX US-09-687-230-1 x AA040621 ..
XX
XX Align seg 1/1 to: AA040621 from: 1 to: 605
XX
XX 120 CTGGGCTGCGCGCGGGGCGGCGGCGGCGGCGGCGGAGATGAGCA 169
XX ||||| ||| |||||::: |||||
XX 1 LeuGlyProAlaAlaGlyAlaA.....MetGlyly 11
XX
XX 170 GAGGACAGAGAGGACAGAGTGGAGC...AAGACCTGACGAGGATAT. 215
XX ||||| ||||| |||||::: |||||
XX 11 slyshIslyshslyshslysalagIutrpArgserSerTyrGluAspTyrA 28
XX
XX 216 .....GTAGAGAGCCCTTGAAGCTGTGCTCTCAAGTAGAGAGG 254
XX ::||| ||||| ||||| ||||| |||||
XX 28 laAspLysProLeuGluLysProLeuLysLeuValLeuYsAlaGlyly 44
XX
XX 255 AAGGAGTCCAGGAGACTCTCCAGGGGAGCTGCGGGGACAGACTCCAGCCT 304
XX ::||| ||||| ||||| |||||
XX 45 serGluValThrgIuLeuSer.....GlyserGlyHIsAspSerSerTy 59
XX

```

```

305 CTTCGAAAGCAAAACGATCAT.....GACAAACACAAAGCAGAAAGC 348
      ::::::::::::::::::::
59 rTyAspAspArgSerAspHisGluArgGluArgHisLysGluLysLysL 76
349 GGAAGGAGGAGGAGGAGGAGGAGGAG.....ATTCCAGGGGAGAA 352
      ::::::::::::::::::::
76 yAluLysLysLysLysSerGluLysGluLysHisLysLeuAspAspGlu 92
393 AAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 442
      ::::::::::::::::::::
93 .....ArgArgLysArgLysGluLysLysLysArgLysArgGluArg 106
443 AACCCGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 489
      ::::::::::::::::::::
106 gEluHisCysAspThrGluGluLysLysLysAspAspProGluLysL 123
490 CTGTGAGATTAGACTTCCTCCGAGAGGAGGAGGAGGAGGAGGAGG 539
      ::::::::::::::::::::
123 yAlaGluValGluProProProAspArgProValArgAlaCysArgThr 139
540 AAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 589
      ::::::::::::::::::::
140 GluProAlaGluLysGluLysSerThrProIleGluGluLysLysL 156
590 GATGAGACAAATTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 639
      ::::::::::::::::::::
156 eLeuArgGluLysGluLysArgLysAspProHisGlyPhePheAlaPhePro 173
640 TGACTGATTTATTTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 689
      ::::::::::::::::::::
173 aLThrAspAlaIleLeuArgProGlyTyrSerMetIleIleLysHisP 189
690 GATTTTATGACATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 739
      ::::::::::::::::::::
190 AspPheGlyThrMetLysAspLysIleValAlaAsnGluTyrLysSer 206
740 AAGAGACTAAAGGATTAATCACTAATGATTAATGATTAATGATTA 789
      ::::::::::::::::::::
206 lThrGluPheLysAlaAspPheLysLeuMetCysAspAsnAlaMetThr 223
790 ACAATTAACACAGAGACCATTTATTTATTAAGTCGCAAGAGGAGG 839
      ::::::::::::::::::::
223 yAsnArgProAspThrValTyrTyrLysLeuAlaLysLysLysLys 239
840 TCAGAGATGAGAAATCTTAGCCAGGAGGAGGAGGAGGAGGAGGAG 872
      ::::::::::::::::::::
240 AlaGlyPheLysMetSerLysGluAlaAlaLeuLysGluLysAsn 256
873 .CAGAGCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 921
      ::::::::::::::::::::
256 pThrAlaValIleGluLysProValProGluValAlaProValGlu 270
922 GAAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 971
      ::::::::::::::::::::
270 ..... 270
972 TGCTGGCAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1021
      ::::::::::::::::::::
270 ..... 270
1022 CAAGAGTCCAGCAAGAAATTAAGAAAGGAGGAGGAGGAGGAGGAG 1071
      ::::::::::::::::::::
271 .ValGluThrAlaLysLysSerLysLysProSerArgGluValIle 287
1072 ATTAGCTTTAA.....AGCAATAT 1091
      ::::::::::::::::::::
287 ySerMetPheGluProGluLysAsnAlaCysSerLeuThrAspSerThr 303
1092 TTAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1141
      ::::::::::::::::::::
304 GluGluHisValLeuAlaLeuValGluHisAlaAlaAspGluAla 320
1142 AAAGTGACGACGAGGCGCTTGTAACAGTCAGTCGCAATTTGAAAG 1191

```

```

      ::::::::::::::::::::
320 pArgIleAsnArgPheLeuProGluGlyLysMetGlyTyrLeuLysArg 337
1192 AACCAATGAGAACACAGACAGCTGGGAGCTTCCTCCATCCTGGAGAT 1241
      ::::::::::::::::::::
337 snGlyAspGlySerLeuLeuTyrSerValAlaAsnThrAlaGluPro 353
1242 GTAGAGAGCCAGGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1291
      ::::::::::::::::::::
354 AlaAspGluGluGluThrHisProValAspLeuSerSerLeuSer 370
1292 ACTTCACTGAGAGTGAATCTTGGACGGGAGGAGGAGGAGGAGGAG 1341
      ::::::::::::::::::::
370 sLeuLeuProGlyPheThrThrLeu...GlyPheLysAspGluArg 386
1342 ACAAGTCATCCAGGTGTATATTTGAATATGAGGAGGAGGAGGAG 1391
      ::::::::::::::::::::
386 snLysValThr.....PheLeuSer...SerAlaThrThrAlaLeu 398
1392 GCACCGCATTTATGACTCCACATTTGCAAAATATGACAAAGATGAT 1441
      ::::::::::::::::::::
399 SerMetGluAsnAsnSerValPheGlyAspLeuLysSerAspGluMet 415
1442 TTTAATCTATTCAACTATGAGGAGGAGGAGGAGGAGGAGGAGGAG 1491
      ::::::::::::::::::::
415 uLeuLeuTyrSerAlaTyrGlyLysArgGluThrGlyValGlnCysAl 432
1492 GCATCCATGATGTTTTGGCCACGCGCAAGATTAATCCGTAATGCA 1541
      ::::::::::::::::::::
432 eLeuGluGluPheValLysAspAlaGlySerTyrSerLysValVal 448
1542 GATAGTTTACTGAGATGTTTAAACAAAGAGGAGGAGGAGGAGGAG 1591
      ::::::::::::::::::::
449 AspAspLeuAspGlnIleThrGlyLysAspHisSerArgThrLeu 465
1592 AGAGATGAGG.....ATGTCATTCCTCGAAGATGAG 1623
      ::::::::::::::::::::
465 eGluLeuLysGluArgArgAsnValProMetLysProProAspGluAl 482
1624 GCCATCTAGACACTT...GACACAGGAGGAGGAGGAGGAGGAGGAG 1670
      ::::::::::::::::::::
482 yAlaGlyAspThrLeuGlyAspSerSerSerSerValLeuGluPheMet 498
1671 GAAGTAGAGCCACAGGCGCTTGACTCCAGTCACTCAAGACAGGCT 1720
      ::::::::::::::::::::
499 SerMetLys.....SerTyrProAspValSerVal 508
1721 AGCGCTGAAAGCAGTAACAAATTTGGCGTTCAGTTGAAGTTTGACT 1770
      ::::::::::::::::::::
508 lAspIleSerMetLeuSerSerLeuGlyLysValLysLysGluLeuAsp 525
1771 CTGAAGAGCTGAATATTCACAGAAAGAACTTGATGAGACACACAGAT 1820
      ::::::::::::::::::::
525 roAspAspSerHisLeu.....AsnLeuAspGluThrThrLysLeu 538
1821 CTGAGGGAATCTCCAGAGACCCAGAAATGAAGTTTGAGACACAGACC 1870
      ::::::::::::::::::::
539 LeuGluAspLeuHisGluAlaGlnAlaGluArgGlyLysSerArgPro 555
1871 TGGAGACATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1901
      ::::::::::::::::::::
555 rSerAsnLeuSerSerLeuSerSerAlaSer 565
seq_name: /SID1/gcgdata/hold-geneseq/genesep-emb1/AA2001.DAT:AA38835
seq_documentation_block:
ID AA38835 standard; Protein: 597 AA.
xx
xx AA38835;
xx
xx
xx 22-Oct-2001 (first entry)
xx
xx Human polypeptide SEQ ID NO 1980.

```


[illegible]

Seq ID	documentation_block
XX	AAU16213 standard; Protein; 199 AA.
XX	AAU16213;
XX	
DT	07-NOV-2001 (first entry)
XX	
DE	Human novel secreted protein, Seq ID 1166.
XX	
KW	Human; immunosuppressive; antiarthritic; antineumatic;
KW	cytostatic; cardiant; vasotropic; cerebroprotective;
KW	neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW	vulnerary; secreted protein; rheumatoid arthritis;
KW	hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW	cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW	nervous system disorder; Alzheimer's disease; infection; ocular disorder
KW	corneal infection; wound healing; epithelial cell proliferation;
KW	skin ageing; food additive; preservative; antiproliferative.
XX	
OS	Homo sapiens.
PN	WO200155322-A2.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01341.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225266.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225268.
PR	14-AUG-2000; 2000US-0225470.
PR	14-AUG-2000; 2000US-0225477.
PR	14-AUG-2000; 2000US-0225757.
PR	14-AUG-2000; 2000US-0225758.
PR	14-AUG-2000; 2000US-0225759.
PR	18-AUG-2000; 2000US-0226279.
PR	22-AUG-2000; 2000US-0226681.
PR	22-AUG-2000; 2000US-0226686.
PR	22-AUG-2000; 2000US-0227182.
PR	23-AUG-2000; 2000US-0227009.
PR	30-AUG-2000; 2000US-0228924.
PR	01-SEP-2000; 2000US-0229287.
PR	01-SEP-2000; 2000US-0229343.
PR	01-SEP-2000; 2000US-0229344.
PR	01-SEP-2000; 2000US-0229345.
PR	05-SEP-2000; 2000US-0229509.
PR	05-SEP-2000; 2000US-0229513.
PR	06-SEP-2000; 2000US-0230437.
PR	06-SEP-2000; 2000US-0230438.
PR	08-SEP-2000; 2000US-0231242.
PR	08-SEP-2000; 2000US-0231243.

310 AAGACAAAACGATCATGACAAACACAGAGAAACCGAAAAAGAGA 359
 |||||
 34 lnsaplysasnaspHisAspLysHisLysaspArgLysArgLysArg 50
 |||||
 360 AAGAAAGGAGAGAGAGATTCAGGGGAGAAAGAGGAGAGAGAGAG 409
 |||||
 51 LysLysGlyGlyLysGlnLeuProGlyGlyGlyLysGlyArgLysArg 67
 |||||
 410 AAGAGTTAAGAGAGATAAAAAGAGAGAGATCGAGACCGGGTGAAGATG 459
 |||||
 67 glnGlyValLys..... 70
 460 AGGACAGAAAAAGATCTCCAGCTGACAGCCCTGTGAGATTAGACTTGCT 509
 |||||
 70 70
 510 CCTGAGAGAGCTCTCACAAGCTCTTAAAGCCAAACAGAAAGATAGACA 559
 |||||
 70 70
 560 GACACCCCTTCAAGAGACTTTGAATCACTGATGAGACAAATTGACAGAGA 609
 |||||
 70 70
 610 AAGATCCAGAGCTCTTCTTTTCATTTCTGTGACTGATTATTGCTCTCT 659
 |||||
 70 70
 660 GGCTACTCCATGATCATTTAAACACCAATGATTTTAGTACCATGAAGA 709
 |||||
 70 70
 710 AAAGATCAAGAACATGACTATGACTCCATAGAGAACTAAAGATTAAGT 759
 |||||
 70 70
 760 TCAACTAATGTTGTTACTAATGCCATGATTACATAAACCAGACACANT 809
 |||||
 70 70
 810 TATTATAAAGTGCMAAGAGCTGTGCTGACTCAGAGATGAAAAATTTCTAG 859
 |||||
 70 70
 860 CCAGGAAGAAATTCAGAGCCCTGAGACAGACATGACTTTCATGCTGACT 909
 |||||
 70 70
 910 TGCAGAAAACCTCGAAAAGCAGAAAGATGAGACAGACCTCAGAGAGTGG 959
 |||||
 70 70
 960 GAGACGAGAGCTCTGCGCAGAGAGAGAGAGACTCTGAGATGCCGA 1009
 |||||
 70 70
 1010 AGCACAGCGCTTCAAGAGTCCACAGAAAGAAATTAAGAAAGAACAAAG 1059
 |||||
 70 70
 1060 ATATGCTTGAAGATTAAGTTTAAACATAATTTAGAGAGAGAGAGAGAG 1109
 |||||
 70 70
 1110 CAGCTTGAAGCGCATGTGAAGAAATCTGAGAGAAAGCTGACAGGCGGCT 1159
 |||||
 70 70
 1160 TGTGAACAGTCAAGTGAATTTGAAGAGAAACAGATGAGAACAGCA 1209
 |||||
 71CysGlnPheGlnArgArgLysProAspLysThrThr 83
 PT

1210 CGTTGGAGCTTCTCATCTGTGATCCATGTGTAGAGAGCCAGGCTAC 1259
 |||||
 83 hrlLeuGlyLeuLeuHisProValAspProIleValGlyGlnProGlyTyr 99
 |||||
 1260 TGGCTGGTGAAGCTGGCATGACAACTGGAAGACTTCAGTCTGAGAGTGA 1309
 |||||
 100 CysProValArgLeuGlyMetThrThrGlyArgLeuGlnSerGlyValAs 116
 |||||
 1310 TACTTGGACGGGCTTCAAGAGATTAAGAGAAAGCAAGTCACTCCAGTGT 1359
 |||||
 116 nhrLeuGlnGlyPheLysGlyLysPheLysArgAsnLysValThrProVal 133
 |||||
 1360 TATATTGAATTAATGGCCCTACAGTCTTATGACAGGATTAATGATGACC 1409
 |||||
 133 eutryLeuAsnTyrGlyProTyrSerSerTyrIleProHisTyrAspSer 149
 |||||
 1410 ACATTTGCAAAATATCAAGAGAGATGATTTGATTTAATCTAATCAACCA 1459
 |||||
 150 ThrPheIleAsnIleSerLysAspAspSerAspLeuIleTyrSerThrTy 166
 |||||
 1460 TGGGGAAGACTCTGATCTTCAAGTGAATTTCAAGCATCCATGACTTTTGG 1509
 |||||
 166 rGlyGlnAspSerAspLeuProSerAspPhe.Ser..... 177
 |||||
 1510 CCAGGTGCCAAGATTATCCGATATGTCATGTCAGATAGTTACTGATGTT 1559
 |||||
 177 177
 1560 TTAACAAAAGAGAGGATTCAGAGACCTTACAAAGATGAGATGTCATT 1609
 |||||
 178H1SC 179
 1610 GCCTGAAGATGAAGCCATACTAGGACTTGACACAGGAAAGAAATGG 1659
 |||||
 179 yseuLysMetLysAlaIleLeuGlyHisLeuThrGlnGlnLysLysTyr 195
 |||||
 1660 AGCAGATTACAGAG 1674
 |||||
 196 ...ArgLeuGlnLys 199
 seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:ABB63028
 seq_documentation_block:
 ID ABB63028 standard; Protein: 861 AA.
 AC ABB63028;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 15876.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 PD 27-SEP-2001.
 PE 23-MAR-2001; 2001MO-US09231.
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 PA (PEKE) PE CORP NY.
 PI Venter JC, Adams M, Li PWD, Myers EW;
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL07131.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 15876; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SO Sequence 861 AA;

alignment_scores:
Quality: 738.50 Length: 824
Ratio: 1.656 Gaps: 24
Percent Similarity: 54.126 Percent Identity: 28.034

alignment_block:
US-09-687-230-1 x ABB63028 ..

Align seg 1/1 to: ABB63028 from: 1 to: 861

168 AAGAACGACAAAGACCAAGTCGGAC...AAGACCTCTACGAGAGTA 214
|||||
5 LysLysHisLysLysAsnLysSerGluArgGluLysTyrGluGluTyr 21
215 T.....GTAGAGAGCCCTTGAGCTGTC 240
|
21 TserGlnHisGlnAspProAlaGlnLeuGlnArgGluLeuLysLeuLeu 38
241 TCMAAGTAGAGGAAAGCAAGTCACCGAAGCTCCACGGGAGCTGGGG 290
|||||
38 euLysValGlySerAsnAlaThrProGluTyrSerAlaAsnSerPromet 54
291 CACGAC.....TCCAGCCCTT 307
55 ValAspGlyLysProProThrAlaAlaGluAlaMetMetSerProValPr 71
308 CGAAGCAAAAGCATCATGACAAACAGAGAGAGAAAGCGGAAAGA 357
|||||
71 ogLugluLeuGlnAspHisGlnGlyHisArgGluArgHisLysLysSer 88
358 GAAAGAGAGAGAGAGAGATTCAGGGGAGAAAGGGGAGAGAAACGG 407
:|||
88 yuLysLysLysLysLys.....LysAspArgGluLysLysHis 100
408 AGAAGAGTTAAGAGAGATAAAGAACCGAGATCGAGACCGGGTGAGAA 457
|||||
101 LysHisHisLysGlu..LysArgHisArgSerArgAspArgHisArgAs 116
458 TGAGGCGAAG.....AAGATCTCCAGTCTC 483
:
116 palagLysSerAspGluAspMetMetAlaGlyAlaAspAspAlaAlaCys 133
484 AC.....
133 erGlyPheAlaProSerSerValAlaProProAlaAlaAspProAspSer 149
485
150 SerGlnAspGlyPheSerPheMetAspAspAspGlnSerGlnProLeuPr 166
485
166 ogLysAsnIleLeuPhePheAlaGlyIleThrThrAspAsnSerProSerA 183

486GCCCTGTGAGATTAGAC..... 503
183 snCyProValThrLysProIleAlaProArgLysLeuAspAlaLeu 199
503
200 MetGlySerSerProAsnSerSerLeuGlnSerSerLeuGlyLe 216
504TGACT..... 509
216 uLleGlySerSerProThrLysProLeuProAspLeuLeuIleProSerP 233
509
233 roSerThrProGluGlyAlaAsnSerLeuAsnAlaLeuThrProLysAla 249
510CCTGAGAGCCCTTCACAGAGCTCTTACGCAAGAGAA... 548
250 LeuGluAlaProLysThrProSerSerSerGlnSerGlnArgGluPr 266
549GAGTAGAACAGACACCCCTTCAGAG 576
266 argSerCyValLeuLysLeuLysGlnGlnLysSerProLeuAsnLysL 283
577 CTTTGAATCAACTGATGAGACAATTGCAGAGAAAGATCCAACTGCTTC 626
|||||
283 euLeuGlnHisLeuLeuArgPheLeuGluLysArgAspProHisGlnPhe 299
627 TTTTCAATTCCTGACTGATTTATGCTCTCTGCTGCTGCTGCTGCTGCT 676
|||||
300 PheAlaThrProValThrAspAspMetAlaProGluLysSerTleI 316
677 TAAACCCCAATGATTTTATGATACATGAAGAAAGATCAAGAACATG 726
|||||
316 eSerArgProMetAspPheSerThrMetArgGlnLysIleAspHisG 333
727 ACTATGCTCCATAGAAGAACTAAGATTAACCTTAATGTACT 776
|||||
333 LuTyrThrAlaLeuThrGluPheThrAspAspPheLysLeuMetCyGlu 349
777 AATGCCATGATTACATTAACCCAGACACCATTTATATTAAGCTGCAAA 826
|||||
350 AsnAlaIleLysTyrAsnHisValAspThrValTyrAsnLysAlaAlaL 366
827 GAAGCTGTGCACCTACAGAAATGAAGATTTAGCCAGAGAAAGATT...C 873
|||||
366 sarGluLeuGlnValGlyMetLysHisLeuGlnProGluAsnLeuMet 383
874 AGAGCTGAAGCAGAGCATAGACTTTCATGCTGCTGCTGCTGCTGCTG 923
:|||
383 rGSerLeuLysProLeuSerGlyTyrMetArgGluLeu.....Thr 396
924 AAGCAGAAAGATGAGACAGACACCTCAGAGTGGGAGAGAGAGCTG 973
397 AlaArgGluLeuGlyPheGluLeuSerSerAsnAsp..... 408
974 CTGGCAGAGAGAGAGAGAGACTGTGAGAT.....G 1005
409 .MetSerArgLysAsnAsnAspSerAlaAspGluGlyAlaSerThrGlyA 425
1006 CCGAAGCAGACAGCCCTTCAGAGTCCACAGAAAGAAATTAAGAAAGAC 1055
|||||
425 laGluGluProArgThrProAlaGlnLeuGlnGluGluArgLysArg 441
1056 AAGATATGCTTGAAGATTAAGTTAAAG.....AA 1087
442 ThrLeuArgLeuGluAsnAlaProLysThrHisPheGluProTyrValAs 458
1088 TAAATTAGAGAGAGAGAG.....GAGCAGCTTGACCGCATCTGAGAG 1131
458 pasPleuThrGlyGluGluIleLeuAlaGlnValGlnAsnAlaAlaGlnG 475

```

1132 AATCTGAGGAAAGCTGACGACGCGCTTGTGAACAGTCAGTCGCAATTT 1181
      ::::::::::::::::::::
475  InAlaLysGlnArGValAsnAlaLysAsnAlaHisLysMetGlyPhe 491
1182 GAAAGAAAGAAACAGATGAGCAACAGCTTGGAGCTTCTCCATCCTGT 1231
      ::::::::::::::::::::
492 LeuArgGlnMetLysAspGlyThrThrThrLeuAsnLeuValIleLysG 508
1232 GGATCCCATTTGTAGAGAGCCAGGCTACTGCTGTGAGACTGGGAAATGA 1281
      ::::::::::::::::::::
508 uGlu.....AsnGlnGlyProGlnArgValValThrIleGlyAspL 522
1282 CAATCGAAGAGCTTCAGTCTGAGAGTGAATACTTTCAGGCGGTCAAGAG 1331
      ::::::::::::::::::::
522 euValGlyLysLeuGlnLysGlySerAlaGlnLeuGlnThrArgGlnVal 538
1332 GATAAAAGAAAGCAAGATCTCAGTCCAGTGTATATGAAATATGAGCCCTA 1381
      ::::::::::::::::::::
539 AspLysArgAsnAlaValAlaArgThrValLysSerLeuAsnTyrGlyAla 555
1382 CAGTCTTATGACCGCATTTATGACTCCACATTTGCAATATACAGCAGG 1431
      ::::::::::::::::::::
555 eAlSerPheAlaProThrPheAspSerArgPheSerThrLeuSerAlaG 572
1432 ATGATTCGATTTATCATCATCAACCTAGGGAGAGACTGTGATCTTCCA 1481
      ::::::::::::::::::::
572 IuGlnThrGlnLeuValAlaLeuArgTyrGlyAspAlaSerSerAlaGlu 588
1482 AGTGATTTACAGCATCATGAGTTTGTGGCCAGTCCCAAGATATTCGCTGA 1531
      ::::::::::::::::::::
589 TyrAlaGlnSerIleLeuGlnPheThrLysAsp...SerAsnTyrGlyTh 604
1532 TGTCATGGCAATAGTTTACTGGATGTTTAAACAAGAGGCGCATTTCCA 1581
      ::::::::::::::::::::
604 rThrIleAlaAsnGlyLeuLeuAspIleLeuThrAsnGlyGlnHisSerL 621
1582 GGACCCCTA.....CAA 1592
      ::::::::::::::::::::
621 ySerLeuAspGlnLeuTyrAsnMetGlnLeuHisSerTyrGlnGlnArg 637
1593 GAGATGAGATGTCATTCCTGAGAGATGAGAGCCATCTAGGACACTTGA 1642
      ::::::::::::::::::::
638 GlnIleGlnLysCysPheGlnGlnGlnGlu.....G 648
1643 CACAGGAAAGAAATGGAGCAGATTTACAGAGTACAGCCACCGGCGCTT 1692
      ::::::::::::::::::::
648 uThrSerSerGlnGlnGlnThrAlaGlnIleGlnGlnGlnGlnGlnGln 665
1693 TGGACTCCAGTCTCAAGACAGGCGTCATAGCGCTGAAAGCAGTACAAT 1742
      ::::::::::::::::::::
665 yTyrLysAsnThrHisValAspPheLysGlyLeuGlnSerLeuGlnGlu 681
1743 TTTCGCGCTTCAGTTGAAATTTTGTACTCTGAGAAAGCTGAAATA.... 1787
      ::::::::::::::::::::
682 LeuGlyIleAspValSerPheLeuAspGlyMetGlnAlaGlnMetLys 698
1788 .....TTCCAGAGAAACTGTATAGACCCACCAAT 1818
      ::::::::::::::::::::
698 rTyrGlnLeuAsnArgArgMetHisGlnHisLysSerGlnAsnLeuThrL 715
1819 TGTCTCAGGAGCTCCAGAGCCCAAGATGAAGCTTTGAGACCCAGCC 1868
      ::::::::::::::::::::
715 euIleGlnLysLeuArgValAlaGlnHisAspArgLeuSer...GlnPro 730
1869 CTTGGGAACATGATCTGTCTCTGGGTCCTCATC...AGAGAAATGCA 1914
      ::::::::::::::::::::
731 LeuProAsnHisLeuGlyLeuValGlnProAlaGlnGlnGlnGlnGln 747
1915 TCTTGTGCAACAAGTACCAATTAATCTTAAGA ATTCCACAGCAAGTAA 1963
      ::::::::::::::::::::
747 nThrAlaGlnGlnLeuThrGlnGlnIleSerAspLeuValLysLysLeuP 764
1964 CTCACAGTGATATGCTAAAGCAGATGAGTTCGAAAAAGCAATGGCGATT 2013

```

```

seq_name: /SIBS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT: AAB95421
seq_documentation_block:
ID AAB95421 standard; Protein: 501 AA.
AC AAB95421;
DT 26-JUN-2001 (first entry)
DE Human protein sequence SEQ ID NO:17823.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
OS Homo sapiens.
PN EP1074617-A2.
PD 07-FEB-2001.
PE 28-JUL-2000; 2000EP-0116126.
PF 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;
XX
WP1: 2001-318749/34.
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs.
XX
PS Claim 8; SEQ ID 17823; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification

```

CC of the present invention.

XX Sequence 501 AA:

alignment_scores:

Quality: 684.00 Length: 473

Ratio: 2.178 Gaps: 11

Percent Similarity: 66.385 Percent Identity: 34.672

alignment_block:

US-09-687-230-1 x AAB95421 ..

Align seg 1/1 to: AAB95421 from: 1 to: 501

```

549 GAAGTAGACAGACAGACCCCTTCAGAAAGCTTTGATCAATGATGAGACA 598
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
19 GlnuSnGluSerThrProIleGlnGlnLeuLeuGlnHisPheLeuArgG1 35
599 ATTGCAGAGAAAGATCCAGAGCTCTTCTTTTCATTTTCCTGACTGATT 648
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
35 nleuGlnArgLysAspProHisGlyPhePheValThrAspA 52
649 TTATTCCTCCGCTACTCCATGATCATTAAACCCCAATGATTTTACT 698
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
52 lalleAlaProGlyTyrSerMetIleIleLysHisPromeCAspPheLy 68
699 ACATGAAAGAAAGATCAAGACATGACTATGATCCATAGAGAAACT 748
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
69 ThrMetLysAspLysIleValAlaSnGluTyrLysSerValThrGluPh 85
749 AAAGCATTAATCAACTAATGTGTACTAATGCCATGATTTACATTAAC 798
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
85 eLyAlaAspPheLysLeuMetCysAspAspAlaMetThrTyrAsnArgP 102
799 CAGAGACATTTATTTAAAGCTGCAAGAGCTTGCTGACTCGAGGATG 848
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
102 roAspThrValTyrTyrLysLeuAlaLysLysIleLeuHisAlaLysPhe 118
849 AAAATTTTACCCAGAAAGATTCAGAGCCTGAGAGACATGACTT 898
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
119 LysMetMetSerLysGluArgLeuLeuAlaLeuLysArgSerMetSerPh 135
899 CATGCTGACTTGCAGAAACTCCAAAGCAGAAAGATGGAACAGACACT 948
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
135 eMetGlnAspMet.....AspPheS 142
949 CACAGAGGGGAGAGAGCGAGCTGCTGCGACAGAGAGAGAGAGACT 998
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
142 eGlnGlnAlaAlaLeuLeuGly.....AsnGlnAspThr 153
999 GGAGATGCCGAAAGCAGACGCTTCAGAGATGCC.....Ag 1033
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
154 AlaValGlnGluProValProGlnValAlaProValGlnValThrAl 170
1034 CAAGAATAATAAAAAGAAAGACAAAGATATGCTTGAAGATTAATAA. 1082
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
170 aLysLysSerLysLysProSerArgGlnValIleSerCysMetPheGluP 187
1083 .....AGCAATTAATTTAGAGAGAGAG 1103
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
187 roGlnGlnAlaLysSerLeuThrAspSerThrAlaGlnGlnHisVal 203
1104 CAGGAGCACTTGACCCGATCGTGAAAGATCTGGAGAAAGACAGCAG 1153
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
204 LeuAlaLeuValGlnHisAlaAlaAspGlnAlaArgAspArgIleAsnArg 220
1154 GCGGCTTGTGAACAGTCAAGTGGAAATTTGAAGAAGAAAGACAGATGAA 1203
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
220 gPheLeuProGlnGlyLysMetGlyTyrLeuLysArgAsnGlnArgLys 237
1204 CAACGACGTTGGGACTTCCATCTGAGATCCATTTGTAGAGAGCCA 1253
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

237 erLeuLeuTyrSerValValAsnThrAlaGlnProAspAlaAspGln 253
1254 GGTACTGCTCGTGAGACTGGGATGACAACTGGAAAGCTTCACTCGG 1303
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 GlnThrHisProValAspLeuSerSerLeuSerLysLeuLeuProG1 270
1304 AGTGAAATCTTTGCGAGGTTCAAGAGAGATAAAGAAAGACAGTCTC 1353
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
270 yPheThrThrLeu...GlyPheLysAspGlnArgArgAsnValThr. 285
1354 CAGTGTATATTTGAATTTATGAGCCCTACAGTCTTTATGACCCGAT 1403
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 .....PheLeuSer...SerAlaThrThrAlaLeuSerMetGlnAsn 298
1404 GACTCCACATTTGCAAAATATCAGCAAGAGATGATTCGATTAATCTAT 1453
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
299 AsnSerValPheGlyAspLeuLysSerAspLysLeuGlnLeuLeuTyrSe 315
1454 AACCTATGGGAGAGACCTGTGATCTTCCAGATGATTTCCAGATGAGT 1503
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
315 lAlaTyrGlnLysAspGlnTyrGlnValGlnCysAlaLeuSerLeuGlnGluP 332
1504 TTTTGGCCAGCTGCCAAGATTTATCCGTATGTCATGAGCAGATAGTTACTG 1553
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
332 heValLysAspAlaGlySerTyrSerLysValValAspAspLeuLeu 348
1554 GATGTTTAAACAAAAGAGGCGATTCAGAGCCTTACAGAGATGAGAG.. 1601
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
349 AspGlnIleThrGlnGlyAspHisSerArgThrLeuPheGlnLeuLysG1 365
1602 .....ATGTCATTCCTGGAAGATGATCAAGCCATACTAGCA 1635
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
365 nArgArgAsnValProMetLysProProAspGlnAlaLysValGlyAspT 382
1636 CACTT...GACACAGAAAGAAATGAGCAGATTTACAGAAAGTACGCA 1682
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 hArgGlyAspSerSerSerSerValLeuGlnPheMetSerMetLys... 397
1683 CGAGGCGTTTGACTCCAGTACTCAGACAGGCTCATAGCGCTGAAGC 1732
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
398 .....SerTyrProAspValSerValAspIleSerMet 408
1733 AGTACAAATTTTGGCGTTCCAGTTGAAGTTTGTACTGCTGAGAAAGCTG 1782
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
408 lLeuSerSerLeuGlnLysValLysLysGlnLeuAspProAspAspSerH 425
1783 AAATATTCAGAAAGAACTTGATGAGACACACAGATTCCTCAGGAGACT 1832
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
425 lSerLys.....AsnLeuAspGlnThrThrLysLeuLeuGlnAspLeu 438
1833 CAGGAAGCCGAGATGAACGTTTGAGCACACAGACCCCTGGAGACATGAT 1882
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
439 HisGlnAlaGlnAlaGlnArgGlnLysSerArgProSerSerAsnLeuSe 455
1883 CTGCTCTTGGGTCCCTCA 1901
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
455 rSerLeuSerAsnAlaSer 461

```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT: AAB41780

seq_documentation_block:

ID AAB41780 standard; protein; 280 AA.

XX AAB41780;

AC AAB41780;

DT 08-FEB-2001 (first entry)

XX Human ORFX ORF1544 polypeptide sequence SEQ ID NO: 3088.

DE Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; KW vulnerable; antiparasitic; antiparkinsonian; neuroprotective; KW anticonvulsant; osteoporotic; antiarthritic; immunosuppressant; cardiant; KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;

263 ValValAsnThrAlaGluProAsnAlaAspGlu 273

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AA95881

seq_documentation_block:

ID AA95881 standard; Protein: 233 AA.

AC AA95881;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:18979.

KW Human; primer: detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

XX 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI: 2001-318749/34.

PS Claim 8; SEQ ID 18979; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AA95881 to

CC AA95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX Sequence 233 AA:

alignment_scores:
Quality: 391.00 Length: 106
Ratio: 4.116 Gaps: 0
Percent Similarity: 89.623 Percent Identity: 65.094

alignment_block:

US-09-687-230-1 x AA95881 ..

Align seg 1/1 to: AA95881 from: 1 to: 233

549 GAAGTAGAACACACACCCCTTCAAGAAAGCTTGAATCACTGATGACACA 598

19 GlnAsnGluSerThrProIleGlnGlnIleuLeuAspHisPheLeuArgG1 35

599 ATTGCACAGAAAGATCCAGTGGCTTCTTTCATTTTCCTGTGACGTAT 648

35 nleuGlnArgLysAspProHisGlyPhePheAlaPheProValIThrAspA 52

649 TTATTTGCTCCTGGCTACTCCATGATCATTTAAACACCCATGGATTTAGT 698

52 IalIeAlaProGlyTyrSerMetIleIleLysHisPrometAspPheGly 68

699 ACCATGAAAGAAAAGATCAAGACATGACTATTCAGTCCATAGAGAACT 748

69 ThrMetLysAspLysIleValAlaAsnGluTyrLysSerValIThrGluPh 85

749 AAAGGATGATCTCAACCTAATGTGTACTAATGCCATGATTTACATTAAC 798

85 elysAlaAspPheLysLeuMetCysAspAsnAlaMetThrTyrAsnArgP 102

799 CAGAGACCATTTATTTATTAAGCTGCAAGAAAGCTGTTGACATCAGAAATG 848

102 roAspThrValTyrTyrLysLeuAlaLysIleLeuHisAlaGlyPhe 118

849 AAAATTTCTAGCCCGAAG 866

119 LysMetMetSerLysGln 124

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAU16626

seq_documentation_block:

ID AAU16626 standard; Protein: 718 AA.

AC AAU16626;

DT 07-NOV-2001 (first entry)

DE Human novel secreted protein, Seq ID 1579.

KW Human; immunosuppressive; antiarthritic; antirheumatic;

KW cytosolic; cardiant; vasotropic; cerebroprotective; nootropic;

KW neuroprotective; antibacterial; virocidic; fungicide; opthalmological;

KW vulnerray; secreted protein; rheumatoid arthritis;

KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;

KW cerebrovascular disorder; cerebral ischemia; angiogenesis;

KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;

KW corneal infection; wound healing; epithelial cell proliferation;

KW skin aging; food additive; preservative; antiproliferative.

OS Homo sapiens.

PN WO200155322-A2.

PD 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01341.

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232387.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 03-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

PR (HUMA-) HUMAN GENOME SCI INC.
PR PI Rosen CA, Barash SC, Ruben SM;
PR PI WPI: 2001-488783/53.
PR DR N-PSDB: AAS26613.
PR DR
PR XX
PR PT New nucleic acid molecules encoding 461 human secreted proteins for
PR PT diagnosing, preventing, treating or ameliorating medical conditions and
PR PT used as food additives or preservatives -
PR XX
PR PS Claim 11; SEQ ID NO 1579; 980bp; English.
PR XX
PR CC The invention relates to isolated nucleic acid molecules and their
PR CC encoded secreted proteins. The nucleic acids and proteins are used to
PR CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
PR CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
PR CC are also used in diagnosing a pathological condition or susceptibility
PR CC to a pathological condition. Antibodies to the proteins can also


```

601 TGCAGAGAAAAGATCCAGTGTCTTTTCATTTCTGCTGACTGATTTT 650
||||| ||||||| : : : : : |||||||
130 enclaspysaspproalaargllephealaglprovalserleuys 146
||||| ||||||| : : : : : |||||||
651 ATTGCCCTGGCTACTCCATGATCATTAACNCCCAATGGATTTTGTAC 700
||||| ||||||| : : : : : |||||||
147 GluValPheAspPryLeuAspHisIleYsHisPheMetAspPheAlaTr 163
||||| ||||||| : : : : : |||||||
701 CATGAAGAAGAAGATCAGAACATGATCTATCATGCTCATGAGAACTAA 750
||||| ||||||| : : : : : |||||||
163 rmetarglysargleuaglualagllytrylsasnleuHisglupheg 180
||||| ||||||| : : : : : |||||||
751 AGCATTACTTCAACTAATGTGTACTAATGCCATGATTTACATTAACCA 800
||||| ||||||| : : : : : |||||||
180 lugalasphaspheaspleuileileaspasnysmetlystryasnalarg 196
||||| ||||||| : : : : : |||||||
801 GAGACCATTTATATTAAGCTGCAGAAAGAGCTTGTGACTCAGGATGAA 850
||||| ||||||| : : : : : |||||||
197 AspThrValPheTyrArgAlaAlaValArgLeuArgAspIndlyglYva 213
||||| ||||||| : : : : : |||||||
851 AATTTTACCCAGGAAGAATTCAGAGCTGGAAGCAGACATGACTTCA 900
||||| ||||||| : : : : : |||||||
213 lValleuArgGlnAlaArg..... 219
||||| ||||||| : : : : : |||||||
901 TGGCTGACTTGCAGAAACTCGAAGACAGAAAGATGAAACAGACACTCA 950
||||| ||||||| : : : : : |||||||
219 ..... 219
||||| ||||||| : : : : : |||||||
951 CAGAGTGGGAGAGAGAGGCTGCTGCGAGAGAGAGAGAGAGCTCTGG 1000
||||| ||||||| : : : : : |||||||
220 .....ArgluValaspserllegl 226
||||| ||||||| : : : : : |||||||
1001 AGATGCCAGAGCA.....CACGCTTCAAGACTCCAGCAAGAA 1041
||||| ||||||| : : : : : |||||||
226 yleuaglulalaserglymeHisleuprogluArgproAlaAlaAlap 243
||||| ||||||| : : : : : |||||||
1042 ATAAAAAG.....AAGACAAAGATRTGCTTGAAGATTAAGTTT 1079
||||| ||||||| : : : : : |||||||
243 roArgTyrProPheSerTrpGluAspValAspArgleuLeuaspProAla 259
||||| ||||||| : : : : : |||||||
1080 AAAAAAGAAATTTTA.....GAGAGAGAGAGAGAGAGAGAGAGCG 1120
||||| ||||||| : : : : : |||||||
260 AsnAlaArgAlaHisleuGlyleuGluGlnleuArgGluLeuLeuasp 276
||||| ||||||| : : : : : |||||||
1121 CATC.....GTGAAGCAATCTGAGAGAAAGCTGACAGC. 1154
||||| ||||||| : : : : : |||||||
276 tleuaspLeuThrCysAlaMetIysSerSerGlySerArgserLyArgA 293
||||| ||||||| : : : : : |||||||
1155 ..CGGCTTGTGAACAGTCAGTGCGAATTTGAAGAAGAAACCAATGGA 1202
||||| ||||||| : : : : : |||||||
293 lAlysleuLeuLysGluIleAlaLeuLeuArgasnLys..... 306
||||| ||||||| : : : : : |||||||
1203 ACACAGACGTGGGACTTCCATCTGTGATGCCATGTGTGAGAGAGCC 1252
||||| ||||||| : : : : : |||||||
307 .....LeuSerGlnGlnHisSer..GlnProLeuProThreGlyP 319
||||| ||||||| : : : : : |||||||
1253 AGGCTACTGCTGCTGTGAGACTGGGAATGACAACTGAGACTTCACTTG 1302
||||| ||||||| : : : : : |||||||
319 ogly..... 320
||||| ||||||| : : : : : |||||||
1303 GAGTGATATCTTTCAGAGGGGTTCAAGAGGAT 1334
||||| ||||||| : : : : : |||||||
321 .....LeuGluGlyPheGluGluasp 327
||||| ||||||| : : : : : |||||||
seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb/AA2001.DAT:AA39231
seq_documentation_block:
ID AAM39231 standard; Protein; 1058 AA.
XX
AC AAM39231;
XX
DF 22-Oct-2001 (first entry)

```

```

XX DE Human polypeptide SEQ ID NO 2376.
XX XX
XX Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX
XX OS Homo sapiens.
XX
XX FN WO200153312-A1.
XX
XX PD 26-JUL-2001.
XX
XX PF 26-DEC-2000; 2000WO-US34263.
XX
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX DR WPI: 2001-442253/47.
XX DR N-PSDB; AAI58387.
XX
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX
XX PS Example 4; SEQ ID NO 2376; 10078bp; English.
XX
XX CC The invention relates to human nucleic acids (AA157798-AA161369) and
XX CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX CC immunosuppressant and cyostatic activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: immune system suppression,
XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening
XX CC C.N.S disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX
XX SO Sequence 1058 AA;

```

alignment_scores:
 Quality: 262.50 Length: 485
 Ratio: 1.156 Gaps: 18
 Percent Similarity: 46.804 Percent Identity: 23.918

alignment_block:
 US-09-687-230-1 x AAM39231 ..

Align seg 1/1 to: AAM39231 from: 1 to: 1058

```

17 CGGCCGGGGCGCGCGCCCTGCTCG..... 46
||||| ||| ||| : : :

```

```

390 ProProGlycysThrArgArgProLeuAsnIleThyGlyAspValGluLeu 406
47 .CGGCGGGGGGTCTCGGGGGCCCCGCTCCGCGCTCGCTCGCTGGCC 95
   ::::::::::::::|
406 tlysAsnGlyValCysArgLys.....GluSerSerValIlysthrValA 421
   ::::::::::::::|
96 GGACCGGAACGGCGCGCACGGCTGGCCCTGGCGGGGGCGGGCAC 145
   ::::::::::::::|
421 rgsSerThrSerLysValArgLysAlaLysAlaLysAlaLysAlaLeu 437
   ::::::::::::::|
146 CGGGGGCCCGGTGGACATGGGCAAGAACAGAACAGACACAGTCGGACA 195
   |||
438 AlagIuProCys..... 441
196 AACACCTCTACGAGAGTATG..... 217
   ::|
442 .AlaValLeuProThrValCysAlaProTyrIleProProGlnArgLeuA 458
   |||
218 .AGAGAGCCCTTGAAGCTGGTCTCAAGTAGGAGG.....GAA 256
   |||
458 snArgIleAlaAsnGlnValAlaIleGlnArgLysGlnPheValGln 474
   |||
257 CGAAGTACCCGA...ACTCTCAAGGGGAGCTGGGGGACGACTCCAGCC 303
   |||
475 ArgAlaIleSerTyrTrpLeuLeuLysArgLeuSerArgAsnGlyAlaPr 491
   |||
304 TCCTCGAAGACAAACATGATGACAAACACAGACAGAAACGGGAAA 353
   |||
491 OleuLeuArgArgLeuGlnSer..SerLeuGlnSerGlnArgSerSerGln 507
   |||
354 AAGAGAAAGAAAGAGAGAGAGATTCACAGGGGAAAGAAAGGAGAAA 403
   |||
508 GlnArgGlnAsnArgLysGlnMetLysAlaLysLysLysLysLys 524
   |||
404 ACGGAGAAAGTTAAGAGAGATAAAGAGAG..... 434
   |||
524 rTrpGlnArgLeuArgHisAspLeuGlnArgAlaArgLeuLeuIleGlu 541
   |||
435 .CGAGATCGAGACCGGGTGAGATGAGCGCAAGAAAGATCTCTCAG 479
   |||
541 euleuArgLysArgGlnLysLeuLysArgGlnValLysValGln 557
   |||
480 TGTCACCCCTGTGAGATTGACTGCTCTCAGAAAGCTCTCAAGAG 529
   |||
558 ValAlaMetGluLeuArgLeu..... 564
530 CTCCTTAGCCAAACAGAGAGTAGAAACAGACACCCCTTCAGAGACT 579
   |||
565 .....ThrProLeuThrValLeuL 571
580 TGAATCAACTGATGAGACAATTGCAGAGAAAGATCCAGTGTCTTCTT 629
   |||
571 euaArgSerValLeuAspGlnLeuGlnAspLysAspProAlaArgIlePhe 587
   |||
630 TCATTTCTCTGACTGATTTTATGCTCCCTGGCCTACATCATTA 679
   |||
588 AlaGlnProValSerLeuLysGlnValProAspTyrLeuAspHisIleLys 604
   |||
680 ACACCAATGGATTTTAGTACATGAAGAAAGAAAGATCAAGAACATGACT 729
   |||
604 snHisProMetSerPheAlaThrMetArgLysArgLeuGlnAlaGlnGly 621
   |||
621 yLysAsnLeuHisGlnPheGlnGlnAspPheAspLeuIleIleAspAsn 637
   |||
730 ATCAGTCCATGAGAAAGAAAGATAACTTCAACATAATGTACTAAT 779
   |||
780 GCCATGATTTTCAATAAACAGACATTTATATTAAGCTGCAAGAA 829
   |||
638 CysMetLysTyrAsnAlaArgAspThrValPheTyrArgAlaAlaValArg 654
   |||
830 GCTGTTCACACTCAGAGAAATGAAGAAATTTTACGCAAGAAATTCAGAGCC 879
   |||
654 gLeuArgAspGlnGlnGlyValValLeuArgGlnAlaArg..... 667

```

```

880 TGAAGACAGACATAGACTTTCATGCTGACTTGCAGAAAACCTCGAAGACAG 929
667 ..... 667
930 AAGATGGAACAGACACACCTTCACAGAGTGGGAGAGACGAGGCTGTGCGCA 979
667 ..... 667
980 GAGAGAGAGAGAGACTCTGAGATGCCGGAAGCA.....CAGCGCT 1020
   |||
668 .ArgGlnValAspSerIleGlyLeuGlnGlnAlaSerClyMetHisLeuP 684
   |||
1021 TCAAGACTCCCGCAAGAAAGAAATTAAGAG.....AAGACAA 1058
   |||
684 rGlnArgProAlaAlaAlaProArgArgProPheSerTrpGluAspVal 700
   |||
1059 GATATGCTTGAGATAGATTAAAGCAATTAATTAA.....GAGAG 1099
   |||
701 AspArgLeuLeuAspProAlaAsnArgAlaHisIleGlyLeuGlnGln 717
   |||
1100 AGAGCAGAGACAGACTTGACCGCATC.....GTGAAAGGAAAT 1134
   |||
717 nLeuArgGlnLeuLeuAspMetLeuAspLeuThrCysAlaMetLysSer 734
   |||
1135 CTGGAGGAAAGCTGACCCAGG...CGGCTTGGAACACTCAGTGGCAATTT 1181
   |||
734 eArgLysSerArgSerLysArgAlaLysLeuLeuLysGlnIleAlaLeu 750
   |||
1182 GAAGAGAAAGAAACCATGATGGAACACAGACGTGGGACTTCTCCATCTGCT 1231
   |||
751 LeuArgAsnLys.....LeuSerGlnGlnHisSer.. 760
1232 GGATCCCATTTGAGAGAGCCAGGCTACTGCTGTGAGACTGGGAATGA 1281
   |||
761 .GlnProLeuProThrGlyProGly..... 768
1282 CAACTGGAAGACTTCACTGTGAGTGAATCTTTGCAGGGGTTCAAGAG 1331
   |||
769 .....LeuGlnGlyPheGlnGln 774
1332 GAT 1334
   |||
775 Asp 775

```

THIS PAGE BLANK (USPTO)

FT ZN_FING 214 264 PHD-TYPE.
 FT DOMAIN 579 649 BROMODOMAIN.
 FT 929 1012 PMP.
 SO SEQUENCE 1058 AA; 119519 MW; 6E7B07E8A030E104 CRC64;

Query Match 8.4%; Score 258.5; DB 1; Length 1058;
 Best Local Similarity 23.5%; Pred. No. 2e-07;
 Matches 109; Conservative 62; Mismatches 136; Indels 157; Gaps 15;

QY 11 DKHLYEYKPKLVKXGNVTELTSGSSGHSLSFEDKNDHKHDKRRKKKKKE 70
 DB 386 DVHTPGCTRRRLNLYGDV-----EMKNGVCKRESSVATVNS-----TSKYKKKKKKAK 434
 QY 71 K-----QIGEEKRRKKRRKKEDKK----- 91
 DB 435 KALAEPCAVLPYVCAPYIPQRLNRIANQVAIORKQFVBRASHYWLKRLSNGAPLAR 494
 QY 92 -----RDRRENEAEKDLQCHAVRLDLPEKPLTSSLAQE-----E 130
 DB 495 RLQSSLSQSRSSQQRNDEEMKAKKELKYWQRLRHLERARLLILKRRKREKREQYK 554
 QY 131 VEG-----TLQELNLMQLQKRDSPAFSPFVTFIAGYSMIIRKHPDEFTME 183
 DB 555 VEGVAMELRITPLTVLRSVLDQLDKDPARITPAQVSLKEVPDYLDHILKHPDFAFMKK 614
 QY 184 KIKNDYOSIEELKDNFKLMCTNAMIYNNRPETIYKAKKLHSGMKILISOERISLQKOS 243
 DB 615 RLEAGQYKNLHFEEDFDLIDNCMKYNAIDYFRAAVRLRQGGVVLRLQAR----- 667
 QY 244 IDPMADLQTRKQKQDGTISQSGEDGCGWOREDESGADA---NAFKSPSKENKK----- 296
 DB 668 -----REVDSIGLEEASGMHLPBPAPAPRRPFSM 697
 QY 297 KDKDMLDEKFSNNL---FREDFOLDRI-----VKESGKLTFR-RLVNSOCERERKPDG 347
 DB 698 EDVDRLLDPAANRAHLERLRLMLDITCMKSSGSRKAKLKEIALRLKK----- 754
 QY 348 TTTGLLHPVDPIVGEPCYCLVRLGTTGRLOSQVNTLOGFKED 391
 DB 755 ---LSQOHS-QPLPTGPG-----LGSFEED 775

RESULT 2
 BRP3_HUMAN STANDARD; PRT: 1214 AA.
 AC 09ULDA;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Bromodomain and PHD finger-containing protein 3 (Fragment).
 GN BRP3 OR KIAA1286.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euteria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20039619; PubMed=10574462;
 RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
 RA Ohara O.;
 RT Prediction of the coding sequences of unidentified human genes. XV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.;
 RL DNA Res. 6:337-345(1999).
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
 CC -1- SIMILARITY: CONTAINS 1 PMP DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AB031112; BAAB6600.1; -;
 DR InterPro: IPR001487; Bromodomain.
 DR InterPro: IPR001965; PHD.
 DR InterPro: IPR000313; PMP.
 DR Pfam: PF00439; bromodomain; 1.
 DR Pfam: PF00628; PHD; 1.
 DR Pfam: PF00855; PMP; 1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00297; BROMO; 1.
 DR SMART: SM00249; PHD; 2.
 DR SMART: SM00293; PMP; 1.
 DR PROSITE: PS00633; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE: PS50014; BROMODOMAIN_2; 1.
 DR PROSITE: PS50812; PMP; 1.
 KW zinc-finger; Bromodomain.
 FT ZN_FING 223 271 PHD-TYPE.
 FT DOMAIN 415 441 GLU-RICH.
 FT 615 685 BROMODOMAIN.
 FT 1085 1168 PMP.
 SO SEQUENCE 1214 AA; 136598 MW; CA490810622109CD CRC64;

Query Match 8.1%; Score 248; DB 1; Length 1214;
 Best Local Similarity 20.8%; Pred. No. 9.2e-07;
 Matches 129; Conservative 95; Mismatches 220; Indels 176; Gaps 20;

QY 22 PLKLVKVGNGNVTLELTSGSSGHSLSFEDKNDHKHDKRRKKKKKGEKQIPGEKGRK 81
 DB 481 PMLAVQIISYLNKLCSTLS-----FQRKNQFMQLHNYWLLKRAQNGVPLRLRLHS 534
 QY 82 RRRVKKKKRDRDRENAEKDLQCHAVRLDLPEKPLTSSLAQEYV--EQPLDEA 139
 DB 535 HLOSRNABQRODEKTSVKKELKYWQRLRHLERARLLILKRRKREKREQYKQDA 594
 QY 140 -----LNQLMRQ-----LQRKDSAFSPFVTFIAGYSMIIRKHPDFTMKETIKN 187
 DB 595 AMELIEMPPNVLRTTLDLQEKDPAHIAEPVNISEVPDYLEFISKPDSTMRKLES 654
 QY 188 NDYOSIEELKDNFKLMCTNAMIYNNRPETIYKAKKLHSGMKILISOERISLQKOSIDFM 247
 DB 655 HLYRTLEFEEDFNLIYTCMKYNAKDTIFHRAAVRLDGLGAILRHARQENIGYD-- 712
 QY 248 ADLQTRKQKQDGTISQSGEDGCGWOREDESGD---AEAHAFKSPSKENKKKKDKDMLD 304
 DB 713 -----PERGTHLPESPKLDFRFSWEDVDNLLIPENRAHLSPEVO----- 753
 QY 305 KFKSNLTEREOLDRI---VKESGKLTFR-RLVNSOCERERKPDGTTTGLLHPVDPIV 361
 DB 754 -----LKLLEKLDLVSMRSGARTRVRLRREINLRK-----LAQPPP-- 797
 QY 362 GEPCYCLVRLGTTGRLOSQVNTLOGFKEDKRNKYTPVLYLNYGYSYAPHYDSTFANI 421
 DB 798 -----QPSLNKTVSNGLPAGQGD--AAVLEQALQE 829
 QY 422 SKDDSLITYSGEDSLPDSIHIFLATCODYRYVADSLIDVITKGHSRTIOEMEM 481
 DB 830 PEDDGG-----RDSKLP-- 860
 QY 482 SLPEDEGHRTLDGKEMQITEVEPPGR----- 860
 DB 861 SPPPP-----PLTKPINDSKPPSKFLPKRVEDELEKSPLOLGNHPTLRLSD 910
 QY 514 STQDRILAKAVTNFQVPEVDFSEAEITQK-----KIDET--TLRLRELQF----- 559
 DB 911 NQINRLSLAPPTAGTPLSGVGRSTVLFKRAKKGVKLQSRPDBVLENGEDHGAAGSPA 970
 QY 560 -----AQNERLSTRPPQNMIC 575

Db 971 SPATSEERHSKRPRSRSC 990

RESULT 3
BREFL_HUMAN STANDARD; PRT; 1214 AA.

AC P55201; O9UHI0;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-OCT-2001 (Rel. 40, Last annotation update)
DE Peregrin (Bromodomain and PHD finger-containing protein 1) (BR140 protein).
GN BRP1 OR BR140.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94161726; PubMed=7906940.
RA Thompson K.A., Wang B., Argaves W.S., Giannotti F.G., Schranck D.P., Rosslahdt E.;
RT "BR140, a novel zinc-finger protein with homology to the TAF250 subunit of TFIID";
RL Biochem. Biophys. Res. Commun. 198:1143-1152(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Hu S.N., Dong W., Zeng Y.X., Yu J., Yang H.M.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: UNKNOWN. POSSIBLE TRANSCRIPTION ACTIVATOR.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: HIGH LEVELS IN TESTIS.
CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC -!- SIMILARITY: CONTAINS 1 PMWP DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M91585; AAB02119.1; .
DR EMBL; AF176815; AAF19605.1; .
DR MIM; 602410; .
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR001965; PHD.
DR InterPro; IPR000313; PMWP.
DR InterPro; IPR000822; Znf-C2H2.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR Pfam; PF00855; PMWP; 1.
DR PRINTS; PRO0503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 2.
DR SMART; SM00293; PMWP; 1.
DR SMART; SM00355; Znf-C2H2; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS50812; PMWP; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein;
KW Zinc-finger; Bromodomain.
FT ZN_FING 21 47 C2H2-TYPE.
FT ZN_FING 273 323 PHD-TYPE.
FT ZN_FING 386 400 PHD-TYPE.
FT DOMAIN 645 715 BROMODOMAIN.
FT DOMAIN 1085 1168 PMWP.
FT CONFLICT 299 299 E -> A (IN REF. 2).

FT CONFLICT 729 729 V -> L (IN REF. 2).
SQ SEQUENCE 1214 AA; 137542 MW; C530CD2F3083A53D CRC64;

Query Match 7.7%; Score 235.5; DB 1; Length 1214;
Best Local Similarity 26.3%; Pred. No. 4.7e-06;
Matches 82; Conservative 58; Mismatches 129; Indels 43; Gaps 10;

QY 82 RRRYEDKKRRDRVENEAE-----KDIQCHAPVRLDLPPEKPYTSSIANQE----- 129
DB 558 RLQTHLQSORNCQVGRSEDDKMWALKEQLSQWRHLRDLERARLLVLEIRREKRE 617
QY 130 -----EVEQPPLOEALNQLMROUQRKDPASFSEFPYTDFTAPCSMIKHHPMFST 180
DB 618 TIKVOITAMEQLTPPLILRLKTLQLEQDKTGNIFFSEPPVLSFVPIYDHIKPMDFFT 677
QY 181 MKEKIKNDYOSIBELKDNFKLMCTNMIYKPPETIYKKAUKLLHSGMKLTISOERQSL 240
DB 678 MKQNLVARYRLNFDFFEDNLIYSNCKLYNADITIRRAVRLREGGAVVQARQAE 737
QY 241 KQSIDFADILQKTRKQKQDGTDSQSGEDGCQWREREDSGDAEAHAFKSPSKENKKRKD 300
DB 738 KMGIDFETGMHIPSILA-GDEATHHTEDAA--EBERLVLLNQKHL---PVEBQLK---- 787
QY 301 MLEDKFSNNIERE---QEQLDRIVKESGKTLRLVNSQCEFERKRPD--GTTTGLLH 355
DB 788 LLERLDEVNASKQSVGRSRAKMKIKEMTALRRKLAH-QREGGRDGPERRGSSRSLSLT 846
QY 356 PVDPVGEPEGYC 367
DB 847 P-----HPAAC 852

RESULT 4
CBP_HUMAN STANDARD; PRT; 2442 AA.

AC 092793; Q16376; Q00147;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE CREB-binding protein.
GN CREBBP OR CBP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97385172; PubMed=9238046.
RA Sobulo O.M., Borrow J., Tomek R., Reshmi S., Harden A., Schlegelberger B., Housman D., Doggett N.A., Rowley J.D., Zelezniak I.E. N.J.;
RT "MLL is fused to CBP, a histone acetyltransferase, in therapy-related acute myeloid leukemia with a t(11;16)(q23;p13.3).";
RT Proc. Natl. Acad. Sci. U.S.A. 94:8732-8737(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97321049; PubMed=9177780;
RA Giles R.H., Petrij F., Dauwerse H.G., den Hollander A.I., Lushnikova T., van Ommen G.J.B., Goodman R.H., Deaven L.L., Doggett N.A., Peters D.J.M., Breuning M.H.;
RT "Construction of a 1.2-Mb contig surrounding, and molecular analysis of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome 16p13.3.";
RT Genomics 42:96-144(1997).
RN [3]
RP SEQUENCE OF 1-405 FROM N.A.
RX MEDLINE=96376968; PubMed=8782817;
RA Borrow J., Stanton V.P., Andersen J.M., Becher R., Behm F.G., Chaganti R.S.K., Clavin C.I., Distche C., Dube I., Frischauf A.M., Housman D., Mittleman F., Volinia S., Watson A.E., Housman D.E.;
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein.";

RL Nat. Genet. 14:33-41(1996).
 CC -1- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
 CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AGUMENTS
 CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
 CC CAMP-RESPONSE GENES.
 CC -1- SUBUNIT: INTERACTS WITH SMAD1, SMAD2 AND SMAD3.
 CC -1- SUBCELLULAR LOCATION: Nucleat.
 CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
 CC T(8;16)(p11;p13) INVOLVING CBP AND MOZ, AND T(11;16)(q23;p13.3)
 CC INVOLVING CBP AND MLL.
 CC -1- DISEASE: DEFECTS IN CREBBP ARE THE CAUSE OF RUBINSTEIN-TAYBI
 CC SYNDROME (RSTS), A DISORDER CHARACTERIZED BY CRANIOFACIAL
 CC ABNORMALITIES, BROAD THUMBS, BROAD BIG TOES, MENTAL RETARDATION
 CC AND A PROPENSITY FOR DEVELOPMENT OF MALIGNANCIES.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 ZN-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U47741; AAC51770.1; -
 DR EMBL: U85962; AAC51331.1; -
 DR EMBL: U89354; AAC51339.1; -
 DR EMBL: U89355; AAC51340.1; -
 DR MIM: 600140; -
 DR InterPro: IPR001487; Bromodomain.
 DR InterPro: IPR003101; KIX.
 DR InterPro: IPR000197; ZNF_finger.
 DR InterPro: IPR000433; ZNF_ZZ.
 DR Pfam: PF00439; bromodomain; 1.
 DR Pfam: PF02172; KIX; 1.
 DR Pfam: PF02135; zf-TAZ; 2.
 DR Pfam: PF00569; zf_1.
 DR PRINTS: PRO0303; BROMODOMAIN.
 DR SMART: SM00297; BROMO; 1.
 DR PROSITE: PS00633; BROMODOMAIN_1; 1.
 DR PROSITE: PS00014; BROMODOMAIN_2; 1.
 DR PROSITE: PS01357; ZF_ZZ_1; 1.
 DR PROSITE: PS01355; ZF_ZZ_2; 1.
 DR Transcription regulation; Nuclear protein; Activator; Bromodomain;
 KW Chromosomal translocation; Zinc-finger.
 FT ZN_FING 1701 1744
 FT DOMAIN 363 430 CYS/HIS-RICH.
 FT DOMAIN 452 683 CREB-BINDING.
 FT DOMAIN 1103 1175 BROMODOMAIN.
 FT DOMAIN 1061 1064 POLY-GLU.
 FT DOMAIN 1199 1487 CYS/HIS-RICH.
 FT DOMAIN 1555 1562 POLY-GLU.
 FT DOMAIN 1675 1849 CYS/HIS-RICH.
 FT DOMAIN 1943 1948 POLY-PRO.
 FT DOMAIN 1967 1970 POLY-GLN.
 FT DOMAIN 2081 2085 POLY-GLN.
 FT DOMAIN 2199 2216 POLY-GLN.
 FT DOMAIN 2245 2248 POLY-GLN.
 FT DOMAIN 2297 2300 POLY-GLN.
 FT CONFLICT 1511 1513 FAE -> NSG (IN REF. 2).
 FT CONFLICT 1724 1725 ED -> VV (IN REF. 2).
 FT CONFLICT 1770 1770 V -> L (IN REF. 2).
 FT CONFLICT 1789 1789 N -> F (IN REF. 2).
 FT CONFLICT 1812 1812 T -> P (IN REF. 2).
 SO SEQUENCE 2442 AA; 265336 MW; 42D084619475F3D2 CRC64;

QY 36 ELSTGSSGHDSLSFEDKND---HDKHKKRKKKKKGKQIPEGKRRRRYKED---- 88
 Db 982 EFNSSQPGPDVPLVLEKKTQAEDEPD-----PGESKGEPSRSEMEEDLQG 1028
 QY 89 ---KKKRDPRVENEAKKDLQCHAPVRLDLPREKPLISLAKQEBVNGT----- 134
 Db 1029 ASQVKEETIAEKKSE-----PMEVD--EKKPEKVVYKEEBSSSSGTJSOSTSPSQ 1079
 QY 135 -----PLOEALNQLMRQLRKDPASAF--FSFVTDPI--APGSMITIKHPMFSTWK 182
 Db 1080 PRKIKPPELRQALPPLTEALYRQPESLPPRPQVDPQLGIPDFYIVKMPDLSTIK 1139
 QY 183 EKIKNDYOSTIEKKNFKLCTNMATYKPEPIYIAKAKKLHSKMLTSRQSLQK 242
 Db 1140 RKLIDGQYDEPQYVDVDMFNMAWLYNRKTSRYKFKPSKL---AEVFDE--IDPVWQ 1194
 QY 243 SLIDFMA-----DLQTKRKQGDGDTSGSGDGC-----WQEREDSGDA 282
 Db 1195 SLGYCCGRKRYEFPQTLCTYQKQLCTIPRDAAYSTQNRHFCEKCFTEIQGENVTLLGD- 1253
 QY 283 EAHAFKSPSEKKNKKDKDMLDEKFSNNLE-----REDEQL----- 318
 Db 1254 -----DPSQPGTTISKQFEKK--KNDTLDPPFPVDCRCKGRKMQLCVLHYDIIMPSEF 1306
 QY 319 ---DRIVKESG-----GKLTTRLVNSQCFERRKPPDGTTLGLHPVDPIV 361
 Db 1307 VCDNCLCKTKGRPRKKNKFSAKRLQTTRLGNHLEDVYKRLRRQN-----HP----- 1352
 QY 362 GEPGYCLVRLGWTGRLSGVNTLQFGKEDKRNKTPVY---YLWGPYSYAPH-YDST 417
 Db 1353 -EAGEYFVRVAVS-----DKTEVAKFGKSRFVDSGKESSEFPYRTKAL 1396
 QY 418 FANISKDSLDLY-----STYGEDSDLPD-----FSIEFLATC---ODYPVYVAD 461
 Db 1397 FAFEEIDGVDFCFGMHVOEGSDCPPWTRRVYISLYDSIHFFPRCLRTAVYHEILLG 1456
 QY 462 SLIDVLTGK---GHSRTLOEMESLPEDEG---HTRLDITGKMEQITVEYPPGRLD- 512
 Db 1457 YLEAYKRLGYVTGHT-----WACPPSGDDYIFCHCPD-----QKTPKRRLOE 1501
 QY 513 -----SSTQDRLIALKAVTNGVVEVDSSEAEIFOKKIDE 549
 Db 1502 WYKKMLDKFAERIIHDYDKIDFKQATEDLISAKELPYF-----EGFWNVLEE 1551
 QY 550 TTRLRLLEQEAQNER 564
 Db 1552 S---IKELQEEER 1563
 RESULT 5
 SN22_HUMAN STANDARD; PRT; 1586 AA.
 AC P51531;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Possible global transcription activator SNF2L2 (SNF2-alpha).
 GN SMARCA2 OR SNF2L2 OR BRM OR SNF2A.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=94038910; PubMed=8223438;
 RA Muehhardt C., Yaniv M.;
 RT "A human homologue of Saccharomyces cerevisiae SNF2/SWI2 and
 RT Drosophila brm genes potentiates transcriptional activation by the
 RT glucocorticoid receptor.";
 RL EMBO J. 12:4279-4290(1993).
 RN [2]
 RP SEQUENCE FROM N.A.

Query Match 6.8%; Score 210.5; DB 1; Length 2442;
 Best Local Similarity 21.2%; Pred. No. 0.0003;
 Matches 143; Conservative 80; Mismatches 213; Indels 239; Gaps 32;


```

RC TISSUE-Brain; PubMed=8208605;
RA MEDLINE=94268902; Kato H.;
RX Chiba H., Muramatsu M., Nomoto A.,
RT "Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and
RT Drosophila brama are transcriptional coactivators cooperating with
RT the estrogen receptor and the retinoic acid receptor."
RL Nucleic Acids Res. 22:1815-1820(1994).
CC -1- FUNCTION: TRANSCRIPTIONAL COACTIVATOR COOPERATING WITH NUCLEAR
CC HORMONE RECEPTORS TO POTENTIATE TRANSCRIPTIONAL ACTIVATION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X72889; CAAS1407.1; -.
DR EMBL: D26155; BAA05142.1; -.
DR MTM: 600014; -.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; SNF2_N.
DR Pfam: PF00439; bromodomain; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00176; SNF2_N; 1.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR PROSITE: PS00633; BROMODOMAIN_1; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
KW Transcription regulation; Nuclear protein; Activator; Bromodomain;
KW ATP-binding; Helicase; Alternative splicing.
FT DOMAIN 216 238 POLY-GLN.
FT DOMAIN 241 249 POLY-GLN.
FT DOMAIN 555 558 POLY-ARG.
FT DOMAIN 639 646 POLY-GLU.
FT NP_BIND 745 752 ATP (POTENTIAL).
FT SITE 847 850 DEAD BOX.
FT DOMAIN 1293 1297 POLY-GLU.
FT DOMAIN 1415 1485 BROMODOMAIN.
FT VARSPPLIC 1397 1414 MISSING (IN SHORT ISOFORM).
FT CONFLICT 239 239 P -> PQOP (IN REF. 2).
FT CONFLICT 390 390 Q -> E (IN REF. 2).
FT CONFLICT 509 509 G -> S (IN REF. 2).
FT CONFLICT 707 707 W -> R (IN REF. 2).
FT CONFLICT 1135 1135 C -> H (IN REF. 2).
FT CONFLICT 1390 1390 D -> V (IN REF. 2).
SQ SEQUENCE 1586 AA; 180762 MW; FA537F2A2392807A CRC64;

Query Match 6.8%; Score 210; DB 1; Length 1586;
Best Local Similarity 24.0%; Pred. No. 0.00018;
Matches 87; Conservative 52; Mismatches 102; Indels 122; Gaps 15;

```

```

QY 156 FSPFVDFIAPGSMIIKHPMDSTMEKIKNDYOSIEELKDNFKIMCTNAMIYKNET 215
DB 1424 IQLPSRREL-PEYELIRKPEVDFEKKIKERINRHSIGDEKDYMLICHNAQTFNEGS 1482
QY 216 IYKAARKLLHSGMKILISGERIOSLKDIPMAILOTRKOKDCTPDSOGEDGCMQRE 275
DB 1483 QIYEDSYLV-----QSV-FKSARQIAKEESEDSDSNEE-----EEE 1519
QY 276 REDSGDAEAFHAFKSPKSNKKKDKMDLEDKRS-----NMLEREQDLRIYKE 324
DB 1520 DEESESSEASVYKIKILNKDDKGRKCKKKRPNKGAKAPVYSDPSDEQDEREGSE 1579
QY 325 SGG 327
DB 1580 GSG 1582

RESULT 6
T2D1_DROME STANDARD; PRT: 2068 AA.
ID T2D1_DROME
AC P51123;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription initiation factor TFIID 230 kDa subunit (TAFTI-230)
DE (TAFTI250) (TBP-associated factor 230 kDa) (P230).
GN TAF250.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 63-75 AND 540-546.
RX MEDLINE=93279463; PubMed=8504928;
RA Kokubo T., Gong D.-W., Yamashita S., Horikoshi M., Reeder R.G.,
RA Nakatani Y.,
RT "Drosophila 230-kD TFIID subunit, a functional homolog of the human
RT cell cycle gene product, negatively regulates DNA binding of the TATA
RL box-binding subunit of TFIID."
RL Genes Dev. 7:1033-1046(1993).
CC -1- FUNCTION: MAY PLAY AN ESSENTIAL ROLE IN TFIID ASSEMBLY BY
CC INTERACTING WITH BOTH TBP AND OTHER TAF. AS WELL AS SERVING TO
CC LINK THE CONTROL OF TRANSCRIPTION TO THE CELL CYCLE. ESSENTIAL FOR
CC PROGRESSION OF THE G1 PHASE OF THE CELL CYCLE. POSSESSES DNA-
CC BINDING ACTIVITY. IS A NEGATIVE REGULATOR OF THE TATA BOX-BINDING
CC ACTIVITY OF TBP.
CC -1- SUBUNIT: TFIID IS COMPOSED OF TBP AND A VARIETY OF TBP-ASSOCIATED
CC FACTORS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC -1- SIMILARITY: CONTAINS 1 HMGB BOX.
CC -1- SIMILARITY: TO HUMAN TAFII-250 (CCG1). SOME TO S. POMBE TAFII-111
CC AND TO S. CEREVISIAE TAFI145.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S61883; AAB26991.2; -.
DR TRANSFAC: T02119; -.
DR FLYBASE: FBgn0010355; TAF250.
DR InterPro: IPR000637; AT hook.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR001878; ZnF_CCHC.
DR Pfam: PF02178; AT_hook; 1.
DR Pfam: PF00439; bromodomain; 2.
DR PRINTS: PR00503; BROMODOMAIN.

```

DR SMART: SM00384; AT_hock; 1.
 DR SMART: SM00297; BROMO; 2.
 DR SMART: SM00343; 2nF_C2HC; 1.
 DR PROSITE: PS00633; BROMODOMAIN 1; 2.
 DR PROSITE: PS50014; BROMODOMAIN 2; 2.
 KW Bromodomain: Nuclear protein; DNA-binding; Cell cycle; Repeat;
 KM Transcription Regulation; Phosphorylation.
 FT DNA_BIND 1247 1360 HMG BOX (POTENTIAL).
 FT DOMAIN 1445 1451 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 1490 1560 BROMODOMAIN 1.
 FT DOMAIN 1612 1682 BROMODOMAIN 2.
 FT DOMAIN 1995 2068 GLN-RICH.
 FT VARIANT 575 575 P -> S.
 SQ SEQUENCE 2068 AA; 232494 MW; AD6A5ABF28B59531 CRC64;

Query Match 6.8%; Score 208; DB 1; Length 2068;
 Best Local Similarity 19.7%; Pred. No. 0.00033;
 Matches 134; Conservative 84; Mismatches 201; Indels 260; Gaps 30;

QY 2 GKKKKHKSDDKILYEE-----YERPLKLVKVG-NEVTELSGSG-----SGHDSL 48
 DB 1321 GSHKKERDSG---YKEVSPSRKKFKLP-DLKLCGCGGVGHMRTNKKACPLYSGMSSL 1376
 QY 49 FEDK-----NDHKKHR-----KK-----KRRGK-----71
 DB 1377 SOSNPLADPDFQSEKEMTDDDLVNDGTVTLSSKILKHGGDDGKRRSGSSGFT 1436
 QY 72 -QIPGEKRRKRRRVEDKKRRDRVENEAEKDLQCHAPVRLDLPPEKPLTSLAKOE 130
 DB 1437 LKVPDAMCKRRV-----GGDLKCDVLQIRNKANKANRRTPV-----1475
 QY 131 VEOTPLQELNOLMROLKRDSPAFSPVTDPIAPGYSMIKHPMDFSTMKERIKNDY 190
 DB 1476 VVLSSTLEIIEHLEIRMPVSP---FLFPVSAKVPDYRVVTKPMDLOTMRREYIQRXY 1532
 QY 191 OSIIEELKDNFKLMCTNAMTYNKPETIYKAAKKLHSGKIIS--QERQSLKQSTDFPA 248
 DB 1533 TSREMFLELDKLVDSILYNGPQSAVTLLAQMSSCELLAEKDKMLREKAINPLL 1592
 QY 249 D-----LQTRKOKDGTDTQSQSGEDGCGWOREDSGDAEHAFAKSPSEK 295
 DB 1593 DDDQVALSFIFDKLHSGKQL-----PEGMPLKP--VAK 1626
 QY 296 KDKDM-----LEDKFSNNLEREQDLDRIVKESGKILTRLVNSQE-----339
 DB 1627 KQVKYTYVYIKRPMDELTGKNIEAHRYSRAEYLA-----DIELIATNCEQYNGSDT 1679
 QY 340 -----FERKKPDGTTGLHPYDPIGEPGYCLVRLGMTTGLQSGVNTLQGFKEKRRK 395
 DB 1680 RTTKFSKILLEYAQTO-----LIEFSEHCGOLENNIAKTO--ERAREN 1720
 QY 396 VMPVLYLNGPYSSAYAPHYSTPANISKDDSLIYSTYGEDSDLPDSFIHFLIATCOPY 455
 DB 1721 -----APERDEAMGN---DDVNF--DRGSRASSPD-----DY 1748
 QY 456 PYVADSLIDVYTKGSHSTTLOEMEKSLPEDEGHTRLDTGKEMQITEVEPPG-----509
 DB 1749 -----IDVGHGSHSSNSIHRSMGAEGSSHTAPA-----VRKADPPGCEVAK 1794
 QY 510 -----RLDSTODRLALAKAVTFGVPEVFDSEEAELIPQKIDETTRLLR 555
 DB 1795 GGRGRKORDPYEEDLQSTDE-----DDDEEDRQ-----1826
 QY 556 ELQEAONERTSTRPPGNNI 574
 DB 1827 EYSEEDENNAASILDGERI 1845

RESULT 7
 SPT7_YEAST STANDARD: PRT; 1332 AA.
 AC P35177;

DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Transcriptional activator SPT7.
 GN SPT7 OR YBR081C OR YBR0739.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C;
 RX MEDLINE=95229044; PubMed=7713415;
 RA Gansheroff L.J., Dollard C., Tan P., Winston F.;
 RT "The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein
 RT important for transcription in vivo."
 RL Genetics 139:523-536(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C;
 RX MEDLINE=95076715; PubMed=7985423;
 RA van der Aart O.J.M., Barthe C., Dolignon F., Aigle M., Crouzet M.,
 RA Steensma H.Y.;
 RT "Sequence analysis of a 31 kb DNA fragment from the right arm of
 RT Saccharomyces cerevisiae chromosome II."
 RL Yeast 10:959-964(1994).
 RN [3]
 RP SEQUENCE OF 1-835 FROM N.A.
 RC STRAIN=5288C;
 RA Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestarazu A.,
 RA Visiers S.;
 RL Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 463-523 FROM N.A.
 RX MEDLINE=92285152; PubMed=1350857;
 RA Haynes S.R., Dollard C., Winston F., Beck S., Trowsdale J.,
 RA Dawd I.B.;
 RT "The bromodomain: a conserved sequence found in human, Drosophila and
 RT yeast proteins."
 RL Nucleic Acids Res. 20:2603-2603(1992).
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF TY ELEMENTS AND POSSIBLY
 CC OTHER GENES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L22537; AAC37424.1; -
 CC EMBL: X76294; CAA53940.1; -
 CC EMBL: Z35950; CAA85026.1; -
 CC EMBL: M87651; AAA5087.1; -
 CC PIR: S41552; S41552.
 CC SGD: S0000285; SPT7.
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; Bromodomain1.
 DR PRINTS: PR00503; Bromodomain.
 DR SMART: SM00297; BROMO; 1.
 DR PROSITE: PS00633; BROMODOMAIN 1; 1.
 DR PROSITE: PS50014; BROMODOMAIN 2; 1.
 KM Transcription Regulation; Nuclear protein; Activator; Bromodomain.
 FT DOMAIN 458 528 BROMODOMAIN.
 SQ SEQUENCE 1332 AA; 152616 MW; 083B6362469244F CRC64;

Query Match 6.5%; Score 199.5; DB 1; Length 1332;
 Best Local Similarity 19.6%; Pred. No. 0.00058;
 Matches 141; Conservative 118; Mismatches 232; Indels 227; Gaps 31;

```

QY 20 EKLPLKLVKGCNEVELSTGSGSHDSLFEDKNDHKRRKRRKGEKQIPG 75
D 275 ERLVLNLSIKETLSKLTNNVEIMGMNKKIYH-SFEYDEKEMIKRLKEESDKMEK 333
QY 76 EERKRRRRVEDKKRRDRENEAEKDLOCHAP-----VRLDLPPE- 118
D 334 GKRRSRSDLAATDEODRENTNDEPTNQKLPPEGSTGSDGNKRRKPSNLDLYNMG 393
QY 119 -----KPLTSLAKQ-----EVEQPTLQELALNOLM 144
D 394 IENSLKHLKLSIQOKSQGLSDYELKHLIMDYRKNKNSKWTSPERIGQELYEACEKV 453
QY 145 KQLOR-KDPSAFSEFPYTDPIAPGYSMITKHPDPSTYKKEIKNNQOSIEELKONFKLM 203
D 454 LELRNYTEHSTPFLNKYSKREAPYHOIILKSMDLNTLVKLKSKFOYDSKQEFVDIMLI 513
QY 204 CTNAMIYKPEPIYK----- 219
D 514 WKNCILTNDSDFHLRGHAIAMOKKSLQILMIRNITIRNADLEKELEDEMEKDKYELD 573
QY 220 -----AAKRLHSGMKILSQERIQ-SLKQSIDFMAD-----LQKTRKQKDG 259
D 574 EEEVAGSGRGLMGMAMLAKEKNGKYSKTVKDEAPFTNDKLSVLPBEGEKEDK 633
QY 260 TDTS-----QSGEDGCGMOREREDSGDAEAAFKSPSE--NKKKDKMLEDK 305
D 634 TASSTVTVHENVNKNEIKENG--KNEQDQVYE-ESSKTEDSSKADAKKDEGLQDK 689
QY 306 FKSNI-----NLERQEQDLRIKESGGLTRIRLYNCOEFERRRPD-----GTTTL 351
D 690 TAEKKEGENCEEEDDEDEDEDE-----WDSQSYLLEKDDRDLEISWVKTYTA 742
QY 352 GLHPVPDIVEPGCYLVRIG-MTTRGLOSQVNTLOGEKEDKRNKVPVLYLNTGYPSSY 410
D 743 KVRAEI-----CLKRTEYFKNKKLNSDEAF--LKNQPKRRFQDLFLKXEQKAL 791
QY 411 APHVDSTFAN-----ISKDSDLI-----YSTYGEIS--DLPSDESIHE--FLAT 451
D 792 ESYRQKTEQNSIMKNGFTYVAKQEDDQLOPHNDHSLNGNAFEKQPDILDLDTFRF-- 849
QY 452 QADPYVWADSLDLVLTFGGHSRTLOEMEMSLPEDEGHTPRLDQKMEQTEVEPPGRL 511
D 850 -QEDD--ISNAIPDIYVGVNTKILDKM-----EDASVDRMLQNC----- 886
QY 512 DSGTODRLIALKAVTNGVPEVDFDSEAEIQQ--KKLDETRLLRELO---EAQNER 564
D 887 -INKQSRLANK---DLGLTPKM--NONITLLOIRHICHKISILRMLOSPLSAQNSR 938

RESULT 8
CBP_MOUSE STANDARD: PRT: 2441 AA.
ID CBP_MOUSE
AC P45481;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE CREB-Binding protein.
GN CREBBP OR CBP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=94019866; PubMed=8413673;
RA Chitliva J.C., Knox R.P.S., Lamb N., Hagihara M., Montminy M.R.,
RA Goodman R.H.;
RT "Phosphorylated CREB binds specifically to the nuclear protein CBP.";
RL Nature 365:855-859(1993).
CC -1- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS

```

```

CC CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
CC CAMP-RESPONSIVE GENES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -1- SIMILARITY: CONTAINS 1 ZINC-FINGER.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
CC EMBL: S66385; AAB28651.1; -.
CC TRANSFAC: T01318; -.
CC MGD: MGI:1098280; Crebbp.
CC InterPro: IPR001487; Bromodomain.
CC InterPro: IPR003101; KIX.
CC InterPro: IPR000197; TAZ_finger.
CC InterPro: IPR000433; ZnF_ZZ.
CC Pfam: PF00439; bromodomain_1.
CC Pfam: PF02172; KIX; 1.
CC Pfam: PF02135; zf-TAZ; 2.
CC Pfam: PF00569; ZZ; 1.
CC PRINTS: PR00503; BROMODOMAIN.
CC SMART: SM00297; BROMO; 1.
CC SMART: SM00291; ZnF_ZZ; 1.
CC PROSITE: PS00633; BROMODOMAIN_1; 1.
CC PROSITE: PS50014; BROMODOMAIN_2; 1.
CC PROSITE: PS01357; ZF_ZZ_1; 1.
CC PROSITE: PS50135; ZF_ZZ_2; 1.
CC KZ Transcription regulation; Nuclear protein; Activator; Bromodomain;
CC zinc_finger.
CC FT DOMAIN 1104 1176 BROMODOMAIN.
CC FT ZN_FING 1702 1745 ZF-ZYPE.
CC FT DOMAIN 1062 1065 POLY-GLU.
CC FT DOMAIN 1556 1563 POLY-GLU.
CC FT DOMAIN 1944 1949 POLY-PRO.
CC FT DOMAIN 1968 1971 POLY-GLN.
CC FT DOMAIN 2082 2086 POLY-GLN.
CC FT DOMAIN 2200 2216 POLY-GLN.
CC FT DOMAIN 2296 2299 POLY-GLN.
CC SQ SEQUENCE 2441 AA; 265474 MW; 0ABB028C3112F419 CRC64;

Query Match 6.4%; Score 198; DB 1; Length 2441;
Best Local Similarity 19.9%; Pred. No. 0.0015;
Matches 139; Conservative 93; Mismatches 219; Indels 248; Gaps 29;

QY 32 NEVELSTGSSGSHDSLFEDKNDHKRRKRRKRRKGEKQIPBEGKRRRRVKED-- 88
D 971 NRVPTPTSTVTSATETSS--QCGPDPVPMLEMKTEVQTDPAEPEPTESKGEPSKMEEDLQ 1028
QY 89 ---KKRRDRENEAEKDLOCHAPVRLDLPPEKPLSLAKQEVQOT----- 134
D 1029 GSSQVKETDTTEQKSEP-----MEVEKKPEVKVEKEEENSSNDTASQSTSPS 1079
QY 135 -----PLQALNOLMROLQKDPASAF-FSEPYTDFI--APGYSMITKHPMDSTM 181
D 1080 QPKRKIRPEELRQALMPTLALTRQDPESLPFQAPVDLPOLGLIPDVFYKKNPMDLSTI 1139
QY 182 KEKIKNDYOSIEELKDNFKLMCTNAMIYKRPETIYKKAARKLHSGMKILSQERIOSLK 241
D 1140 KRLIDTGOYQEPQWYVDVRLFMNMYNKRKTSRVYKFSKSL--AEVEGE-IPVM 1194
QY 242 QSIDFMADLQTRKQKQDTP-----SOSGEGCGM--QKRE 277
D 1195 QSLGCGG---RKYESPQLCYGKQLCTIPDAAYSYQNRHFCGKFTETIGENV 1250
QY 278 DSGDAEAAHAFKSPKSNKKDKMLEDKFSNNLE-----RQEQLE----- 318
D 1251 TLGD-----DPQOPQTITKQDFEKK-KNDITDLPPEFVQCKBCGKRMHICVLHYDII 1302

```

QY 319 -----DRIVEG-----GKTRRLVNSQCEFERKRDGTTTCLLHP 356
 DB 1303 WPSGFWCDLCKTGRPRKRNKESAKRLQTRIGNLIEDRVNLEFRON-----HP 1353
 QY 357 VDPVIGEPCLVRLGTMGTGRLOSGVNTLQGFEDKRNKVPPL-----YLYNGPYSSYAPH 413
 DB 1354 -----EAGEVFRVYVASS-----DKTVEYKPRGKSRFPVDSGEMSEFPY 1392
 QY 414 -----YDSTFANISKDSDLI-----YSTYGEDSDLPDSFSIHEFLAT 451
 DB 1393 RFRALFAPEIDGVDFCFGMHYDFTALLAPHQIOGCVIYSTLD-----SIHFRRR 1444
 QY 452 C-----ODPYVADSLIDVLTGCG-----HSRTLOEME 480
 DB 1445 CLTAVYHEILIGLEYEVKLYVYVTAHWACPPSEGDYIFHCPPDQKIPKRLQEWY 1504
 QY 481 MSLEPEGHTRLDGKEMEQITEVEPPGRIDSSODRLIAKAVNPGVPVPESEEA 540
 DB 1505 KKLMDKAFARLITNDYKDI-----FKQANEDRLTSAKELPY-----EG 1543
 QY 541 EIFOKKIDETTRLLRELQEAONER-----LSTRPPGN 572
 DB 1544 DEWPNVLEES-----IKELQEEERKKEESTASSETPEGS 1579

RESULT 9
 YC40_HUMAN STANDARD; PRT; 733 AA.
 ID YC40_HUMAN
 AC 09JUL10;
 DT 16-OCT-2001 (rel. 40, Created)
 DT 16-OCT-2001 (rel. 40, Last sequence update)
 DT 01-MAR-2002 (rel. 41, Last annotation update)
 DE Hypothetical protein KIAA1240 (Fragment).
 GN KIAA1240.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20039619; PubMed=10574462;
 RA Nagase T., Ishikawa K.-I., Kikuno R., Hirotsawa M., Nomura N.,
 RA Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro."
 RL DNA Res. 6:337-345(1999).
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as the content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AB033066; BAA86554.1; -
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; bromodomain; 1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE; PS0014; BROMODOMAIN_2; 1.
 KW Bromodomain; Hypothetical protein.
 FT NON_TER
 FT DOMAIN 238 320 BROMODOMAIN
 SQ SEQUENCE 733 AA; 83861 MW; 445D3D109D7F817A CRC64;

Query Match 6.3%; Score 195; DB 1; Length 733;

Best Local Similarity 20.3%; Pred. No. 0.00049;
 Matches 112; Conservative 103; Mismatches 204; Indels 134; Gaps 23;
 QY 82 RRRKVEDKKRRDRVENEAE--KDLOCHA-----PVRLLDPEKPLTSSIAKOEY 131
 DB 176 QRIEDRRKRFQELLINQASAPPRKRAALCALMEVLPALPSPRROLSEBKSMEHQ 235
 QY 132 EQTPLOEA--LNQMRQLQKRDPSAFSEFPYDTFAPYSMAIKHPMDFSTPKKIKNN 188
 DB 236 EENTLRELRLFLADYTKRLATDKRFNIFSKPVS DYDE-----VKEPMDLSVTYIKDH 290
 QY 189 DVQSTIELKDNKRLKMTNMIYK-----PETIYKAKKLLHSGMKLL 232
 DB 291 NYLTAPDPLKIDILICSNALFYNPKDKPDGKIIIRHACPLIKDTAAHIIAELDPEFNKIC 350
 QY 233 SQERISLKOSIDFEMD-----LQKT-----RKQDGTPTSQSGEDGCMQRE 275
 DB 351 EELKEKRIKRLGLSVYSEQINPHSTGAKTETVEEAFRRKQKNPMDY-----W--- 398
 QY 276 REDSGDAEHAHAKSPSKRNKKDKDMLIEDKFSNNLEERQDL-----DRIVKES 325
 DB 399 -HNSANKCAFVRKRSRRSSQMGKIIK-KRKYNNLKDEEDTKFADYENHTEDRKLEEN 456
 QY 326 GGLTRRLVNSQCEFERKRDGTTTGLLHPDPIVGEQCYCLVRLGTMGTG-RLOSGVNT 384
 DB 457 G---EFVSTQCHEENGEGETDLSW-----TNDSSCDI-MDLQGGRLNNGAGT 502
 QY 385 LOGFKEDKRNKYTPVLYLNYGYPSYAPHYDSTFANISKDSDLIYSTYGEDSDLPDSFS 444
 DB 503 KENFASTEESSNESLIVASS--SLINPQOTSR-----KETPLKNGCLNGEAS 548
 QY 445 IHEF---LATQODPYVADSLIDVLTGKHSRTLOEMEMSLPDEG---HTRTLDTG 496
 DB 549 TDSFEGIPVLEQ-----NGKLEVVSFCDSGDKSSQKILLEDQSKREKPESTENHG 601
 QY 497 KEMEQITEVEPPGRIDSSODRLIAKAVNPGVPVPESEEAELFOKKLDETRLLRE 556
 DB 602 DDLKLE-----ALECSNNKTL-----EPGSDVEVADALDEKAGAKVKYRKL-- 646
 QY 557 LOEAQNERLSTRP 569
 DB 647 LEQAKTSLIELVP 659
 RESULT 10
 P300_HUMAN STANDARD; PRT; 2414 AA.
 ID P300_HUMAN
 AC 009472;
 DT 15-JUL-1998 (rel. 36, Created)
 DT 15-JUL-1998 (rel. 36, Last sequence update)
 DT 01-MAR-2002 (rel. 41, Last annotation update)
 DE E1A-associated protein p300.
 GN EP300 OR P300.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95011587; PubMed=7523245;
 RA Eckner R., Ewen M.E., Newsome D., Gerdes M., Decaprio J.A.,
 RA Lawrence J.B., Livingston D.M.;
 RT "Molecular cloning and functional analysis of the adenovirus E1A-
 RT associated 300-kD protein (p300) reveals a protein with properties of
 RT a transcriptional adaptor."
 RL Genes Dev. 8:869-884(1994).
 CC -1- FUNCTION: PROBABLE TRANSCRIPTIONAL ADAPTOR REQUIRED FOR THE
 CC ACTIVITY OF CERTAIN COMPLEX TRANSCRIPTIONAL REGULATORY ELEMENTS.
 CC MAY HAVE A FUNCTION IN CELL CYCLE REGULATION. BINDS TO AND MAY BE
 CC INVOLVED IN THE TRANSFORMING CAPACITY OF THE ADENOVIRUS E1A
 CC PROTEIN.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.

Query Match	Similarity	6.2%	Score 190.5	DB 1	Length 2416
Best Local	Similarity 20.4%		Pred No. 0.004		
Matches 130	Conservative 83		Mismatches 249		Indels 175
					Gaps 26

```

OY 395 KATPLVLYNMGPSYVAHDYSTFANISKDSDLIY-----STYGEBSDLPSD----- 442
Db 1354 -----PYRKA-----LFAFEIEDGVCLCFGMHVOEGSDCPENQRRVTSY 1397
OY 443 -FSIHEFLATC-----QDIYVWADSLIDLVLITGGHSRTIQEWMENSLPEDEGTRILLDTCKE 498
Db 1398 LDSVHFEPFKCLRTAVYHEILL-GLEYEVKKLGTY-----TGIWACPPSEG 1443
OY 499 MEQITFVPPRCR-----LD-----SSTQDRLILALKAVTN 527
Db 1444 DDYIFPHCHPPPOCKIPKPKRLQEWYKKMLDKRAVSERIVHDYKIDIFKQATIEDNLTSKELPY 1503
OY 528 FGVPVEVFDSEAEITQKKLDETRTLRLLEQDAQNER 564
Db 1504 F-----EGDFWPNVLEES-----IKELQEEBER 1527

RESULT 11
ATRX_CAEEL STANDARD; PRT; 1359 AA.
ID ATRX_CAEEL
AC Q9U7E0; Q02061;
AD 01-MAR-2002 (Rel. 41, Created)
DE 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DN Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).
GN XNP-1 OR B0041.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=99365296; Pubmed=10433961;
RA Villard L., Fontes M., Ewbank J.D.;
RT "Characterization of xnp-1, a Caenorhabditis elegans gene similar to
RL the human XNP/ATR-X gene."
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN-BRISTOL N2;
RA Fulton R., Wohlmann P.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC GENE EXPRESSION BY AFFECTING CHROMATIN (Potential).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -----
CC This SNF2-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sb.ch/announce/
CC or send an email to license@isb-sb.ch).
CC -----
CC EMBL: AF134186; AAD55361.1; -
CC EMBL: AF000196; AAC24256.1; -
CC WormRep: B0041.7; CE17314.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase.C.
DR InterPro: IPR000330; SNF2_N.
DR Pfam: PF00271; heliase.C; 1.
DR Pfam: PF00176; SNF2_N; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELICC; 1.
KW DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding.
FT NP_BIND 436 503
FT SITE 636 639
FT DOMAIN 67 70 DEAD BOX.
FT DOMAIN 266 272 POLY-ASP.
FT DOMAIN 276 281 POLY-GLU.
FT DOMAIN 372 375 POLY-LYS.
FT DOMAIN 603 608 POLY-LYS.

```



```

QY 7 KHKSDKHLIEEYVEKPLKLVKVGNEVTELTSGSGHDSLFEDKNDHDKHDKRKRK 66
DB 1372 RHREVEYSDSLTEKQWKLKEEGTLEIE-----EYVRKSSKRKRK 1416
QY 67 KKGKQIPEGEKGRKRRRVEDKKRDRVNEAEKDLQCHAPVRLDLPPEKP-LTSSL 125
DB 1417 SDASSPTPTS---TRSDKDESKKQKRRPRAEK-----LSNPRLTR-- 1460
QY 126 AKQEEVQETPLQELNQLMQLQKRDPSAFSPVTDIAGYSMIIRKHPDSTMKRK 185
DB 1461 -KMKIYDAVVIKYKSSGSRQLS-----EVFIQLPSRKEL-PEYVELIRKPVDFKIKERI 1514
QY 186 KKNDOYOSTEELKDNFKLMTCTAMLYNKPEITYYKAAKLHSGKILSQEIOISLKSID 245
DB 1515 RNHHYRSLINDLEKQVMLLCQNAQTFNGLSLEYDS-----IYQSVFTSVRKIE 1565
QY 246 FMADLQKTRKQKDGTDTSQSGEDGCGWQEREDSGDAEHAHAFSPKSNKK--DKDML 303
DB 1566 -----KEDD-----SEGEES-----EHEEGEESGESSESNSVYKIKLGRKKAQ 1606
QY 304 DKFK-----SNLREQQLDR 320
DB 1607 DRLKGRRRRPSGRAPVSDDSSEGEER-DR 1639

RESULT 13
T2D1_HUMAN STANDARD: PRT; 1872 AA.
ID T2D1_HUMAN
AC P21675;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Transcription Initiation factor TFIID 250 kDa subunit (TAFI1-250)
DE (TAFI1250) (TBP-associated factor 250 kDa) (P250) (Cell cycle gene 1
DE protein).
DE TAF2A OR CCG1 OR BAZR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=laryngeal carcinoma;
RX MEDLINE=91246200; PubMed=2038334;
RA Sekiguchi T., Nohiro Y., Nakamura Y., Hisamoto N., Nishimoto T.;
RT "The human CCG1 gene, essential for progression of the G1 phase,
RT encodes a 210-kilodalton nuclear DNA-binding protein.";
RL Mol. Cell. Biol. 11:3317-3325(1991).
RN [2]
RP PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE=89005056; PubMed=3169001;
RA Sekiguchi T., Miyata T., Nishimoto T.;
RT "Molecular cloning of the cDNA of human X chromosomal gene (CCG1)
RT which complements the temperature-sensitive G1 mutants, tsb462 and
RT ts43, of the BHK cell line.";
RL EMBO J. 7:1683-1687(1988).
RN [3]
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
RX MEDLINE=93196704; PubMed=7680771;
RA Ruppert S., Wang E.H., Tjian R.;
RT "Cloning and expression of human TAFII250: a TBP-associated factor
RT implicated in cell cycle regulation.";
RL Nature 362:175-179(1993).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=93196705; PubMed=8450888;
RA Hatake K., Hasegawa S., Takada R., Nakatani Y., Horikoshi M.,
RA Roeder R.G.;
RT "The p250 subunit of native TATA box-binding factor TFIID is the
RT cell-cycle regulatory protein CCG1.";
RL Nature 362:179-181(1993).
CC -!- FUNCTION: MAY PLAY AN ESSENTIAL ROLE IN TFIID ASSEMBLY BY
CC INTERACTING WITH BOTH TBP AND OTHER TAF, AS WELL AS SERVING TO

```

```

CC LINK THE CONTROL OF TRANSCRIPTION TO THE CELL CYCLE. ESSENTIAL
CC FOR PROGRESSION OF THE G1 PHASE OF THE CELL CYCLE. POSSESSES
CC DNA-BINDING ACTIVITY.
CC -!- SUBUNIT: TFIID IS COMPOSED OF TBP AND A VARIETY OF TBP-ASSOCIATED
CC FACTORS.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: PHOSPHORYLATED BY CASEIN KINASE II IN VITRO.
CC -!- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC -!- SIMILARITY: CONTAINS 1 HMG BOX.
CC -!- SIMILARITY: TO DROSOPHILA TAFII-230. SOME TO S.POMBE TAFII-111
CC AND TO S.CEREVISIAE TAFI45.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D90359; BA14374.1; -.
CC EMBL; X07024; CA30073.1; ALT_SEQ.
CC PIR; S03005; S03005.
CC PIR; A40262; A40262.
CC TRASNFRAC; T02206; -.
CC MIM; 313650; -.
CC InterPro; IPR001487; Bromodomain.
CC Pfam; PF00439; bromodomain; 2.
CC PRINTS; PR00503; BROMODOMAIN.
CC SMART; SM00297; BROMO; 2.
CC PROSITE; PS00633; BROMODOMAIN_1; 2.
CC PROSITE; PS50014; BROMODOMAIN_2; 2.
CC KW Bromodomain; Nuclear protein; DNA-binding; Cell cycle; Repeat;
CC Transcription regulation; Phosphorylation.
CC FT DOMAIN 157 165 PRO-RICH.
CC FT DNA_BIND 1195 1273 HMG BOX (POTENTIAL).
CC FT DOMAIN 1351 1358 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT DOMAIN 1397 1467 BROMODOMAIN 1.
CC FT DOMAIN 1520 1590 BROMODOMAIN 2.
CC FT DOMAIN 1627 1872 ASP/GLU-RICH (ACIDIC TAIL).
CC FT SEQUENCE 1872 AA; 212676 MW; 93BE3D181A72ABEB CRC64;

Query Match 5.9%; Score 181.5; DB 1; Length 1872;
Best Local Similarity 21.0%; Pred. No. 0.0093;
Matches 118; Conservative 81; Mismatches 195; Indels 167; Gaps 21;

QY 6 KHKSDKHLIEEYVEKPLKLVKVGNEVTELTSGSGHDSLFEDKNDHDKHDKRKRK 65
DB 1354 KRRRGTVHCDYLNKRPKSHRRRTDPMVTLSTLSEITINDMRDLPNTYPPHPVNAKV 1413
QY 66 KKGKQIPEGEKGRKRRRVEDKKR--DRD-----VNEAEKDLQCHAPV-- 112
DB 1414 VKDYKII---TRPMDLTLENVKRRKRLYPSREERHELELVKNSATYNGPKSHLTGIS 1470
QY 113 ---IDLPEKPLTSSLAQOEVEQETPLQELNQLMQLQKRDPSAF----- 155
DB 1471 QSMIDLDEK-----LKEEKKLALERKAINPL---LDDDDQVAFSFLINDNIYQKMA 1521
QY 156 -----FSPVTDIAGYSMIIRKHPDSTMKRKIKNDVOSIELDNFKLMTNMI 209
DB 1522 VPDSPFPHVPAKKKVPDIYKIVNPMDLETIRKNSHKHTQSRRESFLDDVNLILANSVK 1581
QY 210 YNKPEITYYKAAKLHSGMKILSQ--ERISLQ-----OSIDPMADLQ 251
DB 1582 YNGPESQYTKAQAEIVANCYQITLEYDEHLQLEKDICTAKAEALAELESIDPMPGP 1641
QY 252 KTRKQKDGTDTSQ---GEDGCGWQEREDS--GAEAHAFSPKSNKKDKD----- 300
DB 1642 YTPQPPDIYDNTSLSMSRDASVFODESNMSVLDIPSAITPEKQVQTEDEGDGDLADEE 1701
QY 301 -----MLEDKFSNNLREQQLDRIRYKESGKLTRLVNSQCEFFERRKPDGTTT 350

```

```

Db 1702 GTVQQRQASVLYEDLMS---EGEDDEEDAGSDEEGDN-----PFSAIQLSSEGSDDVG 1753
QY 351 LGLHPVDPVIGEPGYCLVRLGMITGRLQSGVNTLQGFKEKRNKRYPLVY-----401
Db 1754 SGGIRPKOPM-----LQE--NT-----RMDMENESMSMYEGDGGASH 1791
QY 402 -----LNGPYSYAPH--YDSTFANT-----421
Db 1792 GLEDMSNISTSGYEPPKSNTPQDTSPSSIGGYEVSEEEDEDEEQRSGPSVLQVHLSE 1851
QY 422 SKDSDLIYSTYGEDSDLPDSD 442
Db 1852 DEEDSEDHSHNG-DSDDLSD 1871

RESULT 14
BRD2_HUMAN
ID BRD2_HUMAN STANDARD: PRT: 801 AA.
AC P25440:
DT 01-MAY-1992 (Rel. 22, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bromodomain-containing protein 2 (RING3 protein).
GN BRD2 OR RING3 OR KIAA9001.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
ON NCBI_TaxID=9606;
RX MEDLINE=92329974; PubMed=1352711;
RC TISSUE=T-cell;
RA Beck S., Hanson I., Kelly A., Pappin D.J.C., Trowsdale J.;
RT "A homologue of the Drosophila female sterile homeotic (fsh) gene in
RL the class II region of the human MHC.";
RN DNA Seq. 2:203-210(1992).
RP REVISIONS TO N-TERMINUS.
RX MEDLINE=96376536; PubMed=8781126;
RT Thorpe K.L., Abdulla S., Kaufman J., Trowsdale J., Beck S.;
RL "Phylogeny and structure of the RING3 gene.";
RN Immunogenetics 44:391-396(1996).
RP [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,
RA Sato S., Nagase T., Seki T., Ishikawa K.-I., Tabata S.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC -1- SIMILARITY: CONTAINS 1 ET DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62083; AAA43996.1; -
DR EMBL: M80613; AAA68890.1; ALT_INIT.
DR EMBL: D42040; BAA07641.1; -
DR PIR: S18860; S18860.
DR PIR: S40781; S40781.
DR MIM: 601540; -
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; bromodomain.2.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 2.
DR PROSITE: PS00633; BROMODOMAIN_1; 2.
DR PROSITE: PS50014; BROMODOMAIN_2; 2.
KW Bromodomain; Repeat; Nuclear protein.

```

```

FT DOMAIN 91 163 BROMODOMAIN 1.
FT DOMAIN 364 436 BROMODOMAIN 2.
FT DOMAIN 476 515 GLU/SER-RICH.
FT DOMAIN 544 566 ARC/LYS-RICH (HIGHLY BASIC).
FT DOMAIN 775 801 SER-RICH.
FT DOMAIN 555 559 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 638 801 ET DOMAIN.
FT DOMAIN 61 64 POLY-PRO.
FT DOMAIN 492 506 POLY-GLU.
FT DOMAIN 551 559 POLY-LYS.
FT DOMAIN 634 638 POLY-GLU.
FT DOMAIN 775 793 POLY-SER.
SQ SEQUENCE 801 AA; 88060 MW; 9A075EB13507D8E CRC64;

Query Match
Best Local Similarity 27.2%; Score 179.5; DB 1; Length 801;
Matches 67; Conservative 36; Mismatches 90; Indels 53; Gaps 10;

QY 110 PVRDLPPKPLTSSLAKEVEQTPLEAL---NOLMRQLQRKDSAF---FSFPVTF 163
Db 323 PIR---PPRDLDPDQOHOSSKRGKLSQELKHCNGILKELSKHAAVAMPYKPY-DA 378
QY 164 IARG---YSMTIKHPMDFSTMKKIKNNYQSTIELKDNFKLMCTNAMIYKPEITYYA 220
Db 379 SALGLHDYHDIINHPMDLSTVKKRMENRDYRDAQEEPAADVRLAFNSCYKINPDHDVYAM 438
QY 221 AKRL-----LHSGMKILISOERISLQSIDFMADLQKTRQKQDPTDSOS 265
Db 439 ARKLQVFERFRYAKMDEPLFPPPLPVSTAMPGL-----AKSSSESSSES 485
QY 266 GEDGCGQWREREDSGAEHAHAFKSPSENKKDKMDLEKFSNNLERQEQDLRIKES 325
Db 486 SSESSESESESESEDEE-----EESSESDSEEE--RAHRLAEIQEQL-RAVHRQ 533
QY 326 GSKLTR 331
Db 534 LAALSQ 539

RESULT 15
BDP1_YEAST
ID BDP1_YEAST STANDARD: PRT: 686 AA.
AC P35817; O06048;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE BDP1 protein.
GN BDP1 OR YLR399C OR L8084.18.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
ON NCBI_TaxID=4932;
RX MEDLINE=95116323; PubMed=7816623;
RC STRAIN=S288C;
RA Lygenou Z., Conesa C., Lesage P., Swanson R.N., Ruet A., Carlson M.,
RA Sentenac A., Seraphin B.;
RT "The yeast BDP1 gene encodes a transcription factor involved in the
RL expression of a broad class of genes including snRNAs.";
RN Nucleic Acids Res. 22:5332-5340(1994).
RP [2]
RP SEQUENCE FROM N.A.
RA Roeder G.S., Chua P.R.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
RP [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favallo A., Fulton L., Gattung S., Greco T., Kirszen J., Kuuba T.,
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,

```



```

QY 273 QREBDSGDVAHAFKSPBKNKKKDKML---EDFKNSLLEEOOLDRIYESSGKL 322
Db 282 -----AQTHNGRPRKRTIHPKSKDIIYESKKKPKSKRLQDAAMKFCOSVLEL----- 322
QY 330 TRRLVNSOCFEERRKPDGTTTLGLLHPVDPI-VGEPGYC-LVRLGMTGRLQSGVN----- 383
Db 329 -----MAKKHASYNYRPLEPDPVPSMNLPTFYEDYVKRPMDLGTLAKKLNDMOY 376
QY 384 -TLQFKEDKRNKKTPVLYLYLN-----GPISSIAPIH-----YDSTA--NISKDSDL 422
Db 377 QTMDPEEDVY----LVFKNCYFNPDGTLVNMGRLEEVFNKSWADRPMLDDYDSDE 431
QY 429 IYTYGEDSDPLSDFS---IHEFLAT 451
Db 432 .DSRTQGDYDYSEYESSDIDETILIT 457

```

Search completed: July 11, 2002, 15:55:24
Job time: 347 sec

Query Match	5.5%	Score 16.5	DB 1	Length 686
Best Local Similarity	21.3%	Pred. No. 0.016		
Matches 95	Conservative 76	Mismatches 180	Indels 95	Gaps 19

```

QY 35 TELSTGSGHSSSTLPEEDKNDHDKRDKRRKKRKGEOIGEEKGRKKRRRYEDKKRRD 94
D 78 TLANGENGMYAT---GSGAEDBEGQGLKKEGGCGTQ-----EDLENSK 1220
QY 95 DRYENAEKDLQCHAPVRLDLPPEKPLTSSLANOEVEQTPLEALNQLMKLOLRKDP 154
D 121 QELMEVPEK---PAPAP---PPEPPEMNNL-PQNPTRPKHQQKHALLAIVAKRLKARP 177
QY 155 FFSEFPV---TDLIAGYSMLIKHPDESTMEKEIKANNDOYOSIEELKDNFKLMCTNAMY 212
D 173 FLQ-PVDPVKLDIFEFYNYIKRPDLSTIEKRLNVGAVEYEQITDEFNLMVNSIKFNG 231
QY 213 PETTYKAAKLLHSKMLISQERIGLOSIDPMALOKTRKOKDSTDJSQSESDGCM 272
D 232 PNAIGSOMAKRITQASFEKIHNL---MAKDAIPVYIAGRSSAODEAPVIYR----- 281

```

THIS PAGE BLANK (USP)


```

Db 379 --MIGMADRLTS-----OAPVNLVNGPSSAPMIDSWMTWEDTD 421
OY 428 LIYTVGEDSDLPSPDSIEHFLATCODYPYVMAISLDLVLTGKGSRITQEMEMSLPEDE 487
| 422 LFLRFTYGDKSNMNDVSMRFRVGDCEFESEITIG-SLDDLTLDGSEHKTKELE----- 473
OY 488 GHTRLIDTCKEKEQITEVPPGRLDSSDRLIALKAATNFVGVPEVFESEAEIF----- 543
| 474 -----NAGKEVKE--EVNDDEKNETVLSLDDVSSISNLGIEGFLDRIQOVLVPVAV 525
Db 544 -----OKKIDETRLRELQEAQNERLSTRPGNICLLG 578
| 526 ESNINENIPEFMEVNEVHNAVQOOLNHSQKVDLAHQHRLVQOOPPMTMSYQG 580

RESULT 2
12534
hypoetical protein DKFZp434B094.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 02-Sep-2000
C:Accession: T12534
R:Wambutt, R.; Heuener, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A:Reference number: 217524
A:Accession: T12534
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-715 <MAN>
A:Cross-references: EMBL:AL080149
A:Experimental source: adult testis; clone DKFZp434B094
C:Genetics:
A>Note: DKFZp434B094.1
C:Superfamily: bromodomain homology
F:113-168/Domain: bromodomain homology <BRO>

Query Match 8.2%; Score 252.5; DB 2; Length 715;
Best Local Similarity 26.2%; Pred. No. 7.7e-07;
Matches 88; Conservative 53; Mismatches 102; Indels 93; Gaps 11;

OY 86 KEDKKRDRVENEAEKLOCHAPVRLDLPPEKPLTSSIAKOE-----EVEQ----- 133
| 29 QSSSQGREDEEMKAKKELKYWRRLHDLERARLLIELRKREKREQVAYEYVAMEL 88
| 134 --TPLOEALNOLMOLQRRKPSAFSPVTPDIAPGYSMIIRHMPDSTMEKIKNDYQ 191
| 89 RLTPPLTVLLRSVLDQLQDQDPARIFAQPVSLKEVPDYLDHIKHPMDFATMRRLLEAGYK 148
OY 192 STEELKDNFKMLCTNAMYNNKPEITYYKAAKLLHSGMKLISOERQSLQKSIDPMADIQ 251
| 149 NHEFEEDDLIIDCMKYNARDTVFYRAAVRLRDGGVGLRQAR----- 193
OY 252 KTRKQKQDITDSQSGEDGCGWQEREDSGDAE---HAFKSPSKENK---KDKMLMD 304
| 194 -----REVDSIGLEEASGMLPERNAAPRRFRFSDMEDVDRLD 231
OY 305 KRKSNL---EREOLODRI---VKESGKLT/R-LVNSQCEPERRRPDGTTTGLILH 355
| 232 PANRAHLGLEQELRELLDMLDITCAMKSSGSSRSKRLKLEALLRNK-----LSQGH 285
Db 356 PVDPIVGEGYGLVRLGKMTGRLOGSVNLTQFKED 391
| 286 S-QPLPTGPG-----LGGFEED 301

RESULT 3
JC2069
zinc-finger protein, BR140 - human
N:Alternate names: bromodomain protein
C:Species: Homo sapiens (man)
C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 12-Sep-1997
C:Accession: JC2069

```

```

R:Thompson, K.A.; Wang, B.; Argaves, W.S.; Giannotti, F.G.; Schranck, D.P.; Ruoslahti
Biochem. Biophys. Res. Commun. 198, 1143-1152, 1994
A:Title: BR140, a novel zinc-finger protein with homology to the TAF250 subunit of TF
A:Reference number: JC2069; MUID:94161726
A:Accession: JC2069
A:Molecule type: mRNA
A:Residues: 1-1214 <THO>
A:Cross-references: GB:M91585
C:Comment: This is a nuclear protein with broad tissue distribution, but is especially
C:Superfamily: unassigned bromodomain proteins; bromodomain homology
C:Keywords: DNA binding; phosphoprotein; transcription regulation; zinc finger
F:653-708/Domain: bromodomain homology <BRO>
F:23,28,41,44/Binding site: zinc (Cys, Cys, His, His) #status predicted
F:120,205,462/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status
F:276,279,293,296/Binding site: zinc (Cys) #status predicted
F:301,304,317,320/Binding site: zinc (His, Cys, Cys, Cys) #status predicted
F:330,333,350,353/Binding site: zinc (Cys, Cys, His, Cys) #status predicted
F:386,389,401,405/Binding site: zinc (Cys, Cys, His, Cys) #status predicted
F:410,413,444,447/Binding site: zinc (His, Cys, Cys, His) #status predicted

Query Match 7.7%; Score 235.5; DB 2; Length 1214;
Best Local Similarity 26.3%; Pred. No. 1.4e-05;
Matches 82; Conservative 58; Mismatches 129; Indels 43; Gaps 10;

OY 82 RRRVKKKKRDRVENEAE-----KLOCHAPVRLDLPPEKPLTSSIAKOE----- 129
| 558 RLQTHLOSQRNCDQGRSDEKNAKELQKSMQRLHDLERARLLVELIRREKIKRE 617
OY 130 -----EVEQTPLOEALNOLMOLQRRKPSAFSPVTPDIAPGYSMIIRHMPDFT 180
| 618 TIRVQIAMEMOLTPPLILRLKTLLEQDKDNGNITSEVPJSEVPDYLDHIKHPMDF 677
OY 181 MKEKIKNDYOSIEELKDNFKMLCTNAMYNNKPEITYYKAAKLLHSGMKLISOERQSL 240
| 678 MKONLEAYRILNFDDEEDFNILVSNCLAKYNAKDTIFYRAAVRLRDGGAVYQARQME 737
OY 241 KOSIDPMADLOTRKQKQDITDSQSGEDGCGWQEREDSGDAEHAHFKSPSKENKDKD 300
| 738 KMGIDFETGMHIFHSIA-DEATHHTEDAA--EEERLVLENGKH--PVEOLK---- 787
OY 301 MLEDKKSNLRE---EOLODRIKESGKLTIRLVNSQCEPERRRPD--GTTTGLILH 355
| 788 LLERLDEYNASKQSVGRSRRAKMKEMTALRLRLAH--QRETRGDPERHGGSSGLT 846
OY 356 PVDPIVGEGYGC 367
| 847 P-----HPAAC 852
Db 847 P-----HPAAC 852

RESULT 4
D96791
hypoetical protein F15M4.12 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D96791
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hulzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Malt, R.; Matzla
Rizzo, M.; Rooney, T.; Rowley, J.; Sakano, H.
A:Authors: Salzberg, S.L.; Conn, L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: D96791
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-556 <STO>
A:Cross-references: GB:AE005173; NID:g6554481; PIDN:AAPI6663.1; GSPDB:GN00141
C:Genetics:

```

A:Gene: F15M4.12
A:Map position: 1

Query Match 7.3%; Score 225.5; DB 2; Length 556;
Best Local Similarity 23.0%; Pred. No. 1.9e-05;
Matches 108; Conservative 65; Mismatches 168; Indels 129; Gaps 19;

QY 3 KKKKKHSDKLLYEEYKPLKLVKGVNEVELSTGSSGHDSLEEDKN-----53
DB 12 KKKKKRPSLLDQKRALKQOQLQRRNPENEELRSSRNFNFSNRNRNSNSD 71
QY 54 DHDHDKRRK-----KRRKGEKQIPEEGKRRRRYKEDKKRRDRVEN 99
DB 72 DDDRRDKKHLGLNSHEGRDSSNKSQGGDL--DSDANRRKKT-----DSDN 120
QY 100 EAEKDLQCHAFVRLDLPPEKPLTSSLAKEVEEOTPL--QEALQMLQRLKRPSPAFS 157
DB 121 TGER-----ASKATDILQRGSLVESTPLPKLFFIIDRYQKDTYGVYS 167
QY 158 FPPVDFIAPGYSMTIKHPMDFSTMKKIKNNDSIEB---LKDNFKLMCTNAMYKRP 213
DB 168 DPADPEELPDYELIKNPMTFTLRKKIESGAYTTLQFEASLDVF--LICTNAMEYNSA 226
QY 214 ETIYYKAARKLLHSGMKILSGERIOSLQSIDFMADLQTKRKOKDGTDTSGSGEDGCMQ 273
DB 227 DTVYRKQAKMLLAKK-----DFGNLRQESDGEPPVSLSQQPVYK 268
QY 274 REREDSGDAEHAFAFSPSKENKKKDKMLBEDKFSNNLEREOJLDRIYKESGKL----329
DB 269 RGR-----PPGSGLK-----QLEQSLIDTTSISADAAFT 301
QY 330 -----TRLVNSQCFERKRPDGTTLGLHPVDPIVGEPCIV-----RLGMTGRLO 379
DB 302 YAGSSRLSGS--YNNRK--NPPSYGFRAETSAVKNYGMKNVDENRDTYNONSASLO 356
QY 380 -SGVNTLQGFEDKRNKVTYV--LYLNTGPPYSSVAPHYS-----TPANI 421
DB 357 DSSFTLT--LDDMLKQLTPTGLKAEYGYARSLARKYANIGPAWTTANV 403

RESULT 5
S39580
HBRM Protein - human
C/Species: Homo sapiens (man)
C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Sep-1999
C/Accession: S39580
R/Muchardt, C.; Yaniv, M.
EMBL J. 12, 4279-4290, 1993
A>Title: A human homologue of *Saccharomyces cerevisiae* SNF2/SW2 and *Drosophila* brm gene
A:Reference number: S39580; MUID:94038910
A:Accession: S39580
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1586 <MUC>
A:Cross-references: EMBL:X72889; NID:9414116; PIDN:CA51407.1; PID:9414117
F:216-249/Region: glutamine-rich
F:1423-1478/Domain: bromodomain homology <BRO>

Query Match 6.8%; Score 210; DB 2; Length 1586;
Best Local Similarity 24.0%; Pred. No. 0.00054;
Matches 87; Conservative 52; Mismatches 102; Indels 122; Gaps 15;

QY 54 DHDHDKR-----KRRKKKKGEKQIPG-----EER-----GRRK---83
DB 1253 DMDRRREDANRPKRPRIMEDELPSWIKDDAEVERLTCEEEEEKIFFGSGRQRDVYD 1312
QY 84 -----RYKEDKKRR--DRDVENAEKEDLOCHAPVRDLRP 116
DB 1313 SDALTEKQMLRAIEDGNLEMEEEVRLKKRRRRNVDKPAKEDVEK-----AKRRGRP 1367

QY 117 PEKPLTSS---LAKQE-----EEQTPLOEAL---NOLMROLQRKDPSPAF 155
DB 1368 PAEKLSNPPLKTLQKMAIIDTCTYKDKSCVKEKVPNSQLEIRGSSGQLS---EVF 1423
QY 156 FSPVDTFPIAPGYSMTIKHPMDFSTMKKIKNNDSIEBELKDNFKLMCTNAMYKRPET 215
DB 1424 IQDPSRKEL--PEYELIRKPVDFEKIKERIRNHRYSRLGLEDVMLLCHNAQTFNLEGS 1482
QY 216 IYYKAARKLLHSGMKILSGERIOSLQSIDFMADLQTKRKOKDGTDTSGSGEDGCMQ 275
DB 1483 QIYEDSLVL-----QSV--FKSAROKIAKESESDSNEE-----EEE 1519
QY 276 REDSGDAEHAFAFSPSKENKKKDKMLBEDKFS-----NLEREOJLDRIYKE 324
DB 1520 DEESESEAKSVKVIKILNRKDKGKGRKRPNGAKAPVSPDSDSEQDERQSE 1579
QY 325 SGG 327
DB 1580 GSG 1582

RESULT 6
A47371
transcription initiation factor IID 230k chain - fruit fly (*Drosophila melanogaster*)
C/Species: *Drosophila melanogaster*
C/Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
C/Accession: A47371
R/Kokubo, T.; Gong, D.W.; Yamashita, S.; Horikoshi, M.; Roeder, R.G.; Nakatani, Y.
Genes Dev. 7, 1033-1046, 1993
A>Title: *Drosophila* 230-kD TFIID subunit, a functional homolog of the human cell cycl
A:Reference number: A47371; MUID:93279463
A:Accession: A47371
A>Status: preliminary
A:Molecule type: mRNA; protein
A:Residues: 1-2068 <KOK>
A:Cross-references: GB:S61883; NID:9385550; PIDN:AA826991.1; PID:9385551
A>Note: sequence inconsistent with nucleotide translation
A>Note: sequence extracted from NCBI backbone (NCBI:133002, NCBI:133003)
C/Genetics:
A:Gene: FlyBase:Ta1250
A:Cross-references: FlyBase:Fpgn0010355
C/Superfamily: unassigned bromodomain proteins; bromodomain homology
C/Keywords: transcription initiation
F:1498-1553/Domain: bromodomain homology <BRO>
F:1620-1675/Domain: bromodomain homology <BRO>

Query Match 6.8%; Score 208; DB 2; Length 2068;
Best Local Similarity 19.7%; Pred. No. 0.00098;
Matches 134; Conservative 84; Mismatches 201; Indels 260; Gaps 30;

QY 2 GKKKKHSDKLLYEE-----YVEKPLKLVKVG--NEVELSTGS-----SGHDSIL 48
DB 1321 GHSHEKRDSS--YKEVSPRKKFKLP--DLKKGCGAGGVGMIRNRKACPLXSGMSSL 1376
QY 49 FEDEK-----NDHDKRR-----KRR-----KRRKGEK---71
DB 1377 SQSNPSLADDFDQSEKEMTMDDDLNVNDGTQVTLSSKILKRGHGDDGKRRSGSSSGFT 1436
QY 72 -QIPGEKGRKRRRVEDEKKRRDRVENAEKEDLOCHAPVRDLDPPEKPLTSSLAKE 130
DB 1437 LKYPDRAMGKKRRV-----GGDLICDYLQRIHNTANRRRDPV-----1475
QY 131 VEQTPLOEALINQMLQRRKDPSPAFSPVDTFIAPGYSMTIKHPMDFSTMKKIKNNDY 190
DB 1476 VVLSSTILEITHNELRSMPPVSP--FLFPYSAKVPDYRVVTKRPDLDTMREYINGRRY 1532
QY 191 QSLIEELKDNFKLMCTNAMYKRPETIYYKAARKLLHSGMKILS--QERIOSLQSIDFMA 248
DB 1533 TSREMFLEDLKQIVDSNLTLYNGPOSAYTTLAQQWFSSCELLAEKRDLMRLKALNPIL 1592
QY 249 D-----LQTKRKOKDGTDTSGSGEDGCMQREDESDGDAEHAFAFSPSKENK 295

```

Db 1593 DDDQVALSFIFDKLHSQIKQL-----PESWFLKP--VNK 1626
QY 296 KKDQKM-----LEDFKSNLNEREOJLDRIYKESGKILRLVNSOE----- 339
Db 1627 KQYKDYVYIKRPMDELTKGKINIEAHRYSRAVYLA-----DIELVINCQNGSOT 1679
QY 340 ----ERRKPDSTTTLGLHPVDPYIGEPGYCLVRLGKMTTGRLOSGVNTLOGFKEKDRNK 395
Db 1680 RYTKFSKTLFEYAQDQ-----LIERSEHCQGLENNIAKTQ---ERARBN 1720
QY 396 VHPVLYLNGPISSVAPHDSTFANISKDDSLIYSTYGEDSDLPBDFSIHEFLACQY 455
Db 1721 ----APEDEAMGN---DDYFN---DRGSRASSPGD-----DY 1748
QY 456 PYVMADSLDVLTKGSHSPTLOEMKSLPEDGCHTTLTDTGKEMEQITEVEPPG----- 509
Db 1749 ----IDVEGHGCHASSNSIHRSKGAECASSHTA-----VKRPAPGPPEVKR 1794
QY 510 ----RLDSTQDRLALKAQVTFEGVPEVEFDESEAEIIFOKKLDETTLLR 555
Db 1795 GRCRPRKQRPVEDLQCSSTDE-----DDDEEDFPQ----- 1826
QY 556 ELQEAQNERLSTRPPGNMI 574
Db 1827 EYSEDEENNAASITLDGERI 1845

```

```

RESULT 7
S45251
SNP2alpha protein - human
C:Species: Homo sapiens (man)
C:Date: 10-Dec-1994 #sequence_revision 10-Nov-1995 #text_change 28-May-1999
C:Accession: S45251
R:Chida, H.; Muramatsu, M.; Nomoto, A.; Kato, H.
Nucleic Acids Res. 22, 1815-1820, 1994
A:Title: Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and Drosophila brahm
A:Reference number: S45251; MUID:94268902
A:Accession: S45251
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1572 <CH>
A:Cross-references: GB:D61655; NID:9505086; PIDN:BA05142.1; PID:41005684; PID:987661
C:Superfamily: unassigned bromodomain proteins; bromodomain homology
F:1409-1464/Domain: bromodomain homology <BR>

```

```

Query Match 6.7%; Score 205; DB 2; Length 1572;
Best local similarity 23.5%; Pred. No. 0.001;
Matches 82; Conservative 56; Mismatches 99; Indels 112; Gaps 14;

QY 54 DHDKHD---KRRKKRKGKQIQC-----EER-----GKKRR----- 83
Db 1257 DMDRRREDANRPKRPRIMEDELPSMIKDAVERLTCEFEERKIFGSGHQRDVDY 1316
QY 84 ----RYKEDKKR--DRDFENAEKDLQCHAPVRDLP 116
Db 1317 SDALTEKQMLRALEDGNLEMEEEVRLKRRRRNDKPAKADVER-----AKRRGRP 1371
QY 117 PEKPLTSLAKQEEVEQTPLOEALNOLM-ROLOQRKDP-----APSPVPTDFIAPGYS 169
Db 1372 PAEKLSPPNPK-----LTKQNNAIIDIVINYKDSGSGRLSEVFIQLPSRKEL-PEY 1422
QY 170 MIIKHPADSTMEKTKNNNOYOSIEELKONFKLMTNMYNNPPIYYAAAKLHSGM 229
Db 1423 ELIRKPVDFEKIKERIRNNHRSGLDEKVMILCHNAQTFENLGSQIYEDSTVL----- 1477
QY 230 KIISOERIOGLQSGIDFMADLOKTRKQDGTDSOSGEDGCGQORERESGDAEAAHAFKS 289
Db 1478 ----QSV-FKSRQKIAKEESEDSEDSNEE-----EEDDEESESEASVAVK 1519
QY 290 PSKEKKKKDKMLEDKFK-----NNIEREQJLDRIYKESG 327
Db 1520 KIKLNKKDKGRDKGKRRKPRNGKAKAPVVSDFDSEDEDERQSESGS 1568

```

```

RESULT 8
S41552
probable transcription factor SP7 - yeast (Saccharomyces cerevisiae)
M:Alternate names: protein YBR0739; protein YBR081c
C:Species: Saccharomyces cerevisiae
C:Date: 28-Jan-1994 #sequence_revision 09-Sep-1994 #text_change 20-Sep-1999
C:Accession: S41552; S45946; S45948; S40800; S45478; S54985; S59716
R:Gansheroff, L.; Dollard, C.; Tan, P.; Winston, F.
submitted to the EMBL Data Library, July 1993
A:Reference number: S41552
A:Accession: S41552
A:Molecule type: DNA
A:Residues: 1-1332 <GAM>
A:Cross-references: EMBL:L22537; NID:9349189; PIDN:AAC37424.1; PID:9349190
R:Steensma, H.Y.; van der Aart, O.J.M.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45932
A:Accession: S45932
A:Molecule type: DNA
A:Residues: 1-835 <AND>
A:Cross-references: EMBL:Z35950; MIPS:YBR081c
R:Haynes, S.R.; Dollard, C.; Winston, F.; Beck, S.; Trowsdale, J.; David, I.B.
Nucleic Acids Res. 20, 2603, 1992
A:Title: The bromodomain: a conserved sequence found in human, Drosophila and yeast p
A:Reference number: S40800; MUID:92285152
A:Accession: S40800
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 463-523 <HAY>
A:Cross-references: EMBL:M87651; NID:q172683; PIDN:CAA35087.1; PID:q172684
R:van der Aart, O.J.M.; Barthe, C.; Doignon, F.; Aigle, M.; Crouzet, M.; Steensma, H.
Yeast 10, 959-964, 1994
A:Title: Sequence analysis of a 31 kb DNA fragment from the right arm of Saccharomyce
A:Reference number: S45462; MUID:95076715
A:Accession: S45478
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-624/LRGKRRK1,633-1332 <VAN>
A:Cross-references: EMBL:X76294
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1993
R:Gansheroff, L.J.; Dollard, C.; Tan, P.; Winston, F.
Genetics 139, 523-536, 1995
A:Title: The Saccharomyces cerevisiae SP7 gene encodes a very acidic protein importa
A:Reference number: S54985; MUID:95229044
A:Accession: S54985
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1332 <GAM>
A:Cross-references: EMBL:L22537; NID:9349189; PIDN:AAC37424.1; PID:9349190
R:van der Aart, O.J.M.
submitted to the EMBL Data Library, August 1995
A:Reference number: S59702
A:Accession: S59716
A:Molecule type: DNA
A:Residues: 1-1332 <VAN>
A:Cross-references: EMBL:X76294; NID:97424203; PIDN:CAA53940.1; PID:e264674; PID:95583
A:Experimental source: strain S286c
C:Genetics:
A:Gene: SGD:SP17
A:Cross-references: SGD:S0000285; MIPS:YBR081c
A:Map position: 2R
C:Keywords: nucleus; transcription regulation
F:466-521/Domain: bromodomain homology <BR>

```


C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C/Accession: T34036
 R/Fulton, R.; Wohldmann, P.
 A:Title: The EMBL Data Library, April 1997
 A:Description: The sequence of C. elegans cosmid B0041.
 A:Reference number: 221466
 A/Accession: T34036
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-1359 <FNU>
 A/Cross-references: EMBL:AF000196; PIDN:AGC24256.1; GSPDB:GN00019; CESP:B0041.7
 A:Experimental source: strain Bristol N2; clone B0041
 C/Genetics:
 A/Gene: CESP:B0041.7
 A/Map position: 1
 A/Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3

Query Match 6.4%; Score 198; DB 2; Length 1359;
 Best Local Similarity 21.1%; Pred. No. 0.0021;
 Matches 128; Conservative 102; Mismatches 266; Indels 112; Gaps 20;

```

OY 6 KKHSKDHLYEYERPLKLVKGNEVTELTSGSGHDSLFEDKNDHKDKRRKK 65
DB 102 KKSXSKKKVDDKKKKSKK-----KRTSSS-----EDSDSEERKSKKK 143
OY 66 KKKGEQIDGE--EKGRKRRYKEDKKKDRD--RVENEAKDLQCHAPVRLDPEKP 120
DB 144 SKTKKQTSSESESESEERKKSKKNEKSVKRAETSESD-----EDEPKSK 195
OY 121 LNSLAKOEVQOTPLQELNOLMROLDPSAFSPVTPDIAPCYSMITKHPMDFT 180
DB 196 SKKGLAKKKKSSSESESESEDEKE-VKSKSKKKVKKSESEDEADEKKTERKSKS 254
OY 181 MEKIKNDYOSIEELKDNFKLCTNAMYNNRPETIYKAARKL-----HSGMILISQ 234
DB 255 SESSSESEKDEBEKESP-----KPKKKPLAVKKLSDESESESDVAVLQ 304
OY 235 EKIQSLKQSIDMADLOKTRKOKDITISQSGEDGCWQREDSGDADNAHAFSPSKEN 294
DB 305 KKKRGAVTLLISDESEKDKSESEASDVEEKYKKKAKKQSESESGSISIVNRKS 364
OY 295 KKKDKMDLEDK--FKSNLIEREBOOLDRIKESGGKILRRLVNSQCEFERRKPODTTL 351
DB 365 KKKKEPEKKKKIIMDSKLOKJETIDAEAEKER--KKRIEKKKEK----- 409
OY 352 GLIHPVDPYVGEYGLVRLGMTG-----RLQSGVNTLQGEKEDKRNKVTPLYLN-- 403
DB 410 -----NGIVLEEGDITE--MLTGTSSQRLKLSVLDPSSTVDESK-KPEVHNSLV 460
OY 404 --YGYSSYAPH--YDSFANISKDSDLIYSTGEDSDLPDPSIHEFLACODIPIY- 458
DB 461 RLKHAHQHGIQFMTCAFESIDRLDTESSGIIAHCMGLGTLQVITFLHVLVMEKIG 520
OY 459 -MADSLDLVLTGKSHSRLOEMKSLPEDEGHTRLDTGKEMEQLTEVEPPGRLLS--ST 515
DB 521 ECKRVLVAVVPPKNIINNEKQKVLVNDDELDTIDVNE-----LDSYKTI 567
OY 516 ODRILATAA-----VTNGVPEVVDSEAEITFOKKLDETTLLRLLEQAQNERIST 567
DB 568 EDRRALAMWSSKTPSVMIIGYDLFRILITVDDPKKKPKRNRRLEKAKEDPKRYLQN 627
OY 568 RPPGNMTC 575
DB 628 PGPDMVVC 635

```

RESULT 11
 S39161
 CREB-binding protein - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 20-Apr-2000
 C/Accession: S39161

R/Christia, J.C.; Kwok, R.P.S.; Lamb, N.; Hagivara, M.; Montminy, M.R.; Goodman, R.H.
 Nature 365, 855-859, 1993
 A:Title: Phosphorylated CREB binds specifically to the nuclear protein CBP.
 A:Reference number: S39161; MID:94019866
 A/Accession: S39161
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-2441 <CHR>
 A/Cross-references: GB:S66385; NID:q435854; PIDN:AAB28651.1; PID:q435855
 C/Superfamily: unassigned bromodomain proteins; bromodomain homology
 F;1112-1169/Domain: bromodomain homology. <BRO>

Query Match 6.4%; Score 198; DB 2; Length 2441;
 Best Local Similarity 19.9%; Pred. No. 0.0044;
 Matches 139; Conservative 93; Mismatches 219; Indels 248; Gaps 29;

```

OY 32 NEVTELTSGSGHDSLFEDKNDHKDKRRKKRKKGEQIDGEKRRRRYKED-- 88
DB 971 NRVPTSTVSAETSS--QQPGPDVPMLEKMTVEQVDDAEPEPTESGEPSEMEEDLQ 1028
OY 89 ---KKRRDRVENEAKDLQCHAPVRLDPEKPLISLAKOEVQOT----- 134
DB 1029 GSSQVKEETDTTQKSSP-----MEVEKKPEVYKAEKEENSNDTASQSTSPS 1079
OY 135 -----PLQELNOLMROLDPSAF--FSFVTPDI--APGYSMITKHPMDFTW 181
DB 1080 QPKKTIKPEELQALMPLTEALYRODPESLPRQVDDPOLLGIPDIYVAKPMDLSTI 1139
OY 182 KEKIKNDYOSIEELKDNFKLCTNAMYNNRPETIYKAARKLHSGMILISQERISLQ 241
DB 1140 KRLIDGOYQOEPQVYDDVLMFNAMLYNRKTSRYVKCSKL-----AEVFEQ- 1194
OY 242 QSIDFMADLOKTRKOKDGTI-----SQSGEDGCW--QREK 277
DB 1195 QSLGYCGG--RKYFSPQTLCTCYKQLCTIPDAAYSYONRYHFCCKCFETIQENV 1250
OY 278 DSGDAEHAHAFSPSKENKKDKMDLEDKFSNNLE-----REDEQL----- 318
DB 1251 TLGD-----DPSQPTITSKDQFEK-KNDPLDEPEYVDCKEGGRKHQICVLAHYDI 1302
OY 319 -----DRIKESG-----GKILRLVNSQCEFERRKPODTTLGLHP 356
DB 1303 WPSGVCVDCMLKKTGPRKRNKFSARLIQTLRGNLEBRVNFLELRQN-----HP 1353
OY 357 VDPYVGEYGLVRLGMTGRLQSGVNTLQGEKEDKRNKVTPLY--YNTGYSSYAPH 413
DB 1354 -----EAGEVVRVVAAS-----DKTVEYKQMKRFYDSGEMSSFPY 1392
OY 414 -----YSTFANISKDSDLI-----YSTYGEDSDLPDPSIHEFLAT 451
DB 1393 RTKALFAFEIDGVDVCFGMHVQDATALAPHQIQGCVIYSID-----SIHFRRR 1444
OY 452 C-----ODPYVADSLLDVLITGC-----HSRLIQE 480
DB 1445 CLRTAVYHILIGYLEYAKLVVTAHIMACPPSSEDDYIFCHHPDQIKPPKRIQEWY 1504
OY 481 MSLPDEGHTRLDTGKEMEQLTEVEPPGRLLDSQDRILATAKAVYNGVPEVVDSEEA 540
DB 1505 KMLDKAPAFERILINDYKOT-----FKQANDRLITSKELPYF-----EG 1543
OY 541 EIFQKKLDETTLLRLLEQAQNER-----LSTRPPGN 572
DB 1544 DFWPNVLEES--IKELQEEERKKEESTASSETPEGS 1579

```

RESULT 12
 S39162
 Transcription coactivator CREB-binding protein - human
 N/Alternate names: CBP; RSTs; Rubinstein-Taybi syndrome (RNS) protein
 C/Species: Homo sapiens (man)
 C/Date: 07-Oct-1994 #sequence_revision 17-Nov-1995 #text_change 10-Dec-1999
 C/Accession: S39162; S60345; I58096

R:Chivla, J.C.; Kwok, R.P.S.; Lamb, N.; Hagihara, M.; Montminy, M.R.; Goodman, R.H.
Nature 365, 855-859, 1993
A:Title: Phosphorylated CREB binds specifically to the nuclear protein CBP.
A:Reference number: S39161; MUID:94019866
A:Accession: S39162
A:Molecule type: mRNA
A:Residues: 1-2440 <CHR>
A:Note: differences with the mouse sequence are shown
R:Lundblad, J.R.; Kwok, R.P.S.; Laurence, M.E.; Harter, M.L.; Goodman, R.H.
Nature 374, 85-88, 1995
A:Title: Adenoviral E1A-associated protein p300 as a functional homologue of the transcr
A:Reference number: S60344; MUID:95174889
A:Accession: S60345
A:Status: preliminary
A:Molecule type: protein
R:Perlt, F.; Giles, R.H.; Dauwerse, H.G.; Sails, J.J.; Hennekam, R.C.; Masuno, M.; Tom
Nature 376, 348-351, 1995
A:Title: Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activat
A:Reference number: 158096; MUID:95356817
A:Accession: 158096
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 352-356 <PERT>
A:Cross-References: GB:089354; NID:q1041931; GB:S78936; NID:q1041931
A:Note: this translation is not annotated in Genbank entry S78936, release 112.0
A:Note: this sequence with a termination mutation is from a patient with Rubinstein-Tayb
C:Genetics:
A:Gene: GDB:CREBP; RTS: CBP; RSNs
A:Cross-References: GDB:471159; OMIM:180849; OMIM:600140
A:Map position: 16p13.3-16p13.3
A:Note: defects in this gene may result in Rubinstein-Taybi syndrome
C:Superfamily: unassigned bromodomain proteins; bromodomain homology
C:Keywords: phosphoprotein; transcription; zinc finger
F:462-661/Domain: CREB binding #status predicted <CBR>
F:1111-1168/Domain: bromodomain homology <BRO>
F:1283-1311/Region: zinc finger CCCC motif
F:1707-1732/Region: zinc finger CCCC motif
F:78,781,745,1112,1533,2063,2354/Binding site: phosphate (Ser) (covalent) (by calmodulin
F:1771/Binding site: phosphate (Ser) (covalent) (by protein kinase A) #status predicted

Query Match 6.2%; Score 192; DB 2; Length 2440;
Best Local Similarity 19.7%; Pred. No. 0.0097;
Matches 137; Conservative 87; Mismatches 216; Indels 254; Gaps 29;

36 ELTSSGSHSSLFEDKND--HDHKHRRKKRKKGEKQIPGEKGRKRRRYE-- 88
Db 982 ETSNQQGCPVYVLEMTETGAEDETPD-----PGSKGEPREEMHEEDLOG 1028
Qy 89 -KKRRDRVENEAEKDLQCHAPVRLDLPPEKPLTSSSLAKQEEVEQT----- 134
Db 1029 ASQVKEETDLAEQKSEP-----MEVEDKKPEVKEVEEESSSNGTASOSTPSQ 1079
Qy 135 -----PLQELNQLMQLORRKSAP--FSFPTVDF--ARGYSMLIKHPDFSMK 182
Db 1080 PRKIKPEELRQALMPTLEALYRDPESLPFROPVDPQLGIDPYDIYKPMDLSTIK 1139
Qy 183 EKIKNDYQSIIEELDKNFILCTNAMYKPKETIYYKAKKILHSGMKILSOERISLQK 242
Db 1140 RKLDTGQYQERWQYVDDVWLMFNNAWLYNRTSRYKFCSTL-----AVFQGE-IDPYWQ 1194
Qy 243 SIDFWA-----DLQTRKQKDGTDTSQSGEDGCC-----WQREKSDSDA 282
Db 1195 SLGCGCGKRYEFPQTLCCYCKQCLTIPRDAAYYSYONRYHFCCEKCTETIAGENVTLD- 1253
Qy 283 EAHAFKSPKSKENKKKDKOMLEDKFKSNLLE-----REGRQL----- 318
Db 1254 -----DPSQPOQTITSKQFEKK--KNDTLDEPPYVDCKEGCKRMQICVLHYDITWPGF 1306
Qy 319 --DRIVEESG-----GKLTIRLVNSQCEFFERRKRPDGTTLGLHPEVDIV 361
Db 1307 VCDNCKTKTGPRKKNKSAKRLQTRTIRGNLHEDRVNKNFLRRQN-----HD----- 1352

Qy 362 GEPGYCLVRLGTTGRLQSGVNTLQGEKDKRNKVPVL---YLNTPYSSIVAPH----- 413
Db 1353 -EAGEVFVRVAVASS-----DKTEVVRPGMKSRFVDSGSESESPYRTKAL 1396
Qy 414 -----YSTFANISKDSDLL-----YSTYGEISDLPSPDSIHEPLATC----- 452
Db 1397 FAFEEIDGVVCFPGMHDVDTALAPHOIGCVIYSLD-----SHFFPRCLRTA 1448
Qy 453 -----QDPYVADSLDLVLTGK-----HSRTLQEMENSLPE 485
Db 1449 VYHEILLIGYEVYKLVYTAHITWACPPSGDDYIFCHPDPDKIPKPKRLQENYKMKMD 1508
Qy 486 DEGHTRTLDGKMEQITVEEPGRULDSYQDRLLAKAVTNFGVPEVVDSEAEITFOK 545
Db 1509 KAFERIRIINYKDI-----FKQANEDRLTSAKEPLPYF-----EGDEWPN 1547
Qy 546 KLDETTLRLLEQEAQNER-----LSTNPPGN 572
Db 1548 VLEBS--IKELQDEERKKEESTAASETPEGS 1578

RESULT 13
T40984
transcription factor btf1 homolog SPC1450.02 - fission yeast (Schizosaccharomyces po
N/Alternate names: protein SPC191.13
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 08-Dec-2000
A:Accession: T40984; T41225
R/Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL data library, March 1999
A:Reference number: 221962
A:Accession: T40984
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-578 <LYN1>
A:Cross-References: EMBL:AL049559; NID:q4581510; PIDN:CA840169.1; PID:q4581512; GSPDB
R/Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL data library, March 1998
A:Reference number: 221904
A:Accession: T41225
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-111 <LYN2>
A:Cross-References: EMBL:AL049644; NID:q4678667; PIDN:CA841059.1; PID:q4678679; GSPDB
A:Experimental source: strain 972h-; cosmid c191
C:Genetics: <C145>
A:Gene: SPDB:SPC1450.02
A:Map position: 3
C:Genetics: <C191>
A:Gene: SPDB:SPC191.13
A:Map position: 1
C:Superfamily: bromodomain homology
F:109-166/Domain: bromodomain homology <BRO1>
F:279-336/Domain: bromodomain homology <BRO2>

Query Match 6.2%; Score 191.5; DB 2; Length 578;
Best Local Similarity 21.3%; Pred. No. 0.0017;
Matches 123; Conservative 77; Mismatches 209; Indels 169; Gaps 24;

32 NEVTELSSGSHSS--LFEDKN--DHDHKHRRKKRKKGEKQIPGEKGRKRRRYE 87
Db 8 NEVKAETDEIANDSSPOLNGDNNGSSDGHNDENESLSR--KRSSGATYVGDLLQOEKE 66
Qy 88 DKRRDRVENEAEKDLQCHAPVRLDLPPEKPLTSSSLAKQEEVYQFPLQELNQLMQL 147
Db 67 SMPKKEPEPYKKINGS-----GMP-----PQAKCLAIIVQL 100
Qy 148 QKRDSPAFSPVTFD--IAPGYSMTIKHPMDFSTMEKIKNDYQSIIEELDKNFILMCT 205
Db 101 KTKNSAPFKVYVDPIDKONIDPYPIYVKNPMDLGTETIKLTSYEVSVQEFIDMNLMS 160

QY 206 NAMIYKPEITYYKAAKLLHSGMKILSOERISLQSIDFMADLOKTRKQDGTDSQS 265
 Db 161 NCFLYNTESPVSGMKAL-----QEVFERQLQQLPDAGPAAPAKKSKASTAP-- 213
 QY 266 GEGGCGHQREREDSGAEAAFAFKSPSKENKKDKMLDEKFKSNLREOEDLDRIKES 325
 Db 214 -----PRTRRNSVSTSTASVAASAPAKSPAVLPE----- 245
 QY 326 GGLTRRLVNSQCEFERRRKPDGTTLLGLH-----PVDPVIGE-PGYC-LVR 370
 Db 246 -GPRRRKRNNSQMF-----CSTVLKELKRYOESAFPFYQVDPVADCDPDYVIK 298
 QY 371 LGMTTGRLQSGVN-----TLQGFKEKDKRNVTPVLYLNTGYSSYADPHYDSTANISKD 425
 Db 299 EPMDLSTIOSKLNKNEYSTLEEFESD-----ILLM-----FNNCF----- 333
 QY 426 SDLIYSYGGSDSLPSPFSHFELATCODPYPMADSLDLVLKGGHSLRLQEMKSLPE 465
 Db 334 -----TYN-----PQTPVH-----VMGRLQLEVENFEKWEARPKFDPATLVKQ 371
 QY 486 DEGHTRTL-DTGKEMEOI-TEVEPPG-----RLDSTQDRLIALKA----- 524
 Db 372 QEAETDALPDNGEERELKMSSEELNGAKFAVADKQISMLODTLEAKKAKKMNMRPRR 431
 QY 525 --VTNGCVVEVVDSEAE-----IFOKKIDETTRLIRE 556
 Db 432 DLTREYGPITYAMQNELAERCNYLSAEOLSNVAELIRE 469

RESULT 14

T28145
 C:Species: Gallus gallus (chicken)
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T28145
 R:Milne, S.; Kaufman, J.; Beck, S.
 submitted to the EMBL Data Library, May 1998
 A:Description: DNA sequencing and analysis of the chicken major histocompatibility comp
 A:Reference number: Z20475
 A:Accession: T28145
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-733 <MIL>
 A:Cross-references: EMBL:AL023516; NID:e1292539; PID:e1292549; PIDN:CA118965.1
 A:Experimental source: clone CB12
 C:Genetics:
 A:Gene: RING3
 A:Map position: 16
 A:Introns: 64/3; 110/3; 158/1; 227/3; 351/3; 394/3; 479/3; 546/2; 650/1; 691/1
 C:Superfamily: unassigned bromodomain proteins; bromodomain homology
 F:52-109/Domain: bromodomain homology <BR01>
 F:323-380/Domain: bromodomain homology <BR02>

Query Match 6.2%; Score 191; DB 2: Length 733;

Best Local Similarity 28.5%; Pred. No. 0.0024;

Matches 71; Conservative 43; Mismatches 103; Indels 32; Gaps 11;

QY 110 PYRLDLPPEKPLTSSLANOEVEQTPLOEAL---NQLMKLOLRKPSAF---FSPVPTDF 163
 Db 275 PTK---PKKDLDPDSOQHTS-KKGLKLSBQLKYCGILLKELSKHAAVAMFYKPV-DA 329
 QY 164 IAPG---TSMIKHPMDSTMAKEKTKNDYOSIEELKDNFKLMCTNMIYKPEITYYKA 220
 Db 330 SALGLHADVHEIIKHPMDLSTIKRKEMNRDHYDAQEFADVRLMFNCYKPNPDHVVAM 389
 QY 221 AKKLL-----HSGMKILSOERISLQK-SIDFMADLOKTRKQDGTDSQSGDGCQWRE 275
 Db 390 ARKLDQVDFEFSAKMPDPQDASPPSVAPLIGALSKSSSESSDDEDED-----D 443
 QY 276 REDSGDAHAHAFKSPSKENKKDKMLDEKFKSNLREOEDLDRIKESGGLTRRLRVN 335

Db 444 EDDDEDESSSSSSSDSPSSDSEEE-----RANRLAELOQL-RAVHEQALALSQGPVS 496
 QY 336 SQCEFERRK 344
 Db 497 KPRKKRREK 505

RESULT 15

A54277
 A:transcription adaptor protein p300 - human
 C:Species: Homo sapiens (man)
 C:Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 10-Dec-1999
 C:Accession: A54277; S60344
 R:Ekner, R.; Ewen, M.E.; Newsome, D.; Geres, M.; Decaprio, J.A.; Lawrence, J.B.; Li
 Genes Dev. 8, 869-884, 1994
 A>Title: Molecular cloning and functional analysis of the adenovirus E1A-associated 3
 A:Reference number: A54277; MID:95011587
 A:Accession: A54277
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2414 <ECK>
 A:Cross-references: GB:U01877; NID:9495300; PIDN:AA118639.1; PID:9495301
 A>Note: In the authors' translation 941-Ser is shown after 961 and consequently, resi
 R:Lundblad, J.R.; Kwok, R.P.S.; Laurance, M.E.; Harter, M.L.; Goodman, R.H.
 Nature 374, 85-88, 1995
 A>Title: Adenoviral E1A-associated protein p300 as a functional homologue of the tran
 A:Reference number: S60344; MID:95174889
 A:Accession: S60344
 A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 552-660 <LUN>
 C:Genetics:
 A:Gene: GDB:EP300
 A:Cross-references: GDB:9862958; OMIM:502700
 A:Map position: 22q13.2-22q13.2
 C:Superfamily: unassigned bromodomain proteins; bromodomain homology
 C:Keywords: phosphoprotein; transcription; zinc finger
 F:1075-1132/Domain: bromodomain homology <BR0>
 F:89,507,1136,1295,1497,1834,1977,2062,2320/Binding site: phosphate (Ser) (covalent)
 F:1734/Binding site: phosphate (Ser) (covalent) (by protein kinase A) #status predict

Query Match 6.2%; Score 190.5; DB 2: Length 2414;

Best Local Similarity 20.4%; Pred. No. 0.012;

Matches 130; Conservative 83; Mismatches 249; Indels 175; Gaps 26;

QY 38 STGSSGHDSSLFEDKNDHDKKRRKKRKGKQJGFEKGRKRRVEDKKRRDRV 97
 Db 956 STSSTVNSQALAEK---QPSQEVKKEAKMEVDQPPADTQPDISEKVEDECKMESTET 1012
 QY 98 ENAEKDLQCHAPVRLDLPPEKPLTSSLA---KQEVQTPLOEALNOLMKLOLRKPS 153
 Db 1013 E-ERSTELTEIKEEDQSTGTSATGSSPAPGSKKIFPEELRLQMLPTLEATVRODE 1071
 QY 154 AP-FSPVPTDFI-AGYMIILKHPMDSTMAKEKTKNDYOSIEELKDNFKLMCTNMIY 210
 Db 1072 SLPEKQVPDQQLGLPDYDIKSPMDLSTIKRKLDGTQGPWPQVDDIWMFNAMLY 1131
 QY 211 NKPETIYYKAAKLLHSGMKILSOERISLQSIDMA-----DLQTRKQ 256
 Db 1132 NKTSTVYVYCKL-----SEVEQF-IDPVMOISLIGCCGRKLEFSPTQLCCYQKQLCTIP 1186
 QY 257 KDGTDTSQSGEDGCG-----WQEREDSGAEAAFAFKSPSKEN-KKKDKMLDEKFK 307
 Db 1187 RDAYYSYQNRHYHCEKCEFNIEGSESVSLGDDPQSPQPTTINKEQSKRKNDLDELPE 1246
 QY 308 SNMLREDEQL-----DRVKESEG-----GKLTIRLV 334
 Db 1247 CTEGCRKMQLQICVLHHEIIPAGFVDCGLKRSARTRENKFSAKRLDPTSLGTLENV 1306
 QY 335 NSQCEFERRKPDGTTLLGLHVPDPIYGEPGYCLRLGTTGRTGSLQSGVTLQGFEDKRN 394
 Db 1307 NDLRLQNHPEGSEYTVRVVHASDKTVE-----VKPGKAKAFVDSG-EMAESEF----- 1353

```

QY 395 KVPVLYLNGPYSSYAPHNDSTFANISKDDSLIY-----STYGEDSDLPD----- 442
Db 1354 -----PYRKA-----LFAFEIDGVDLCFGGMHVQEGSDCPPNORRYISY 1397
QY 443 -FSIHEFIATC---QDYPVVADSLDVLTKGHSRTLQEMEMSLPDEGHTRTLDTGKE 498
Db 1398 LDSVHFRPKCLRTAVYHEILI-GYLEYVKKLGTY-----TGHIMACPPSEG 1443
QY 499 MEQITEVEPPGR-----LD-----SSTODRLIALKAVTN 527
Db 1444 DDYIFHCHPPDQKIPKPKRLQEWYKKMLDKAVSERIVHDYKDIFKQATEDRLTSAKELPY 1503
QY 528 FGVPVEVPDSEAEIFQKKIDETRLRLBELOEAONER 564
Db 1504 F-----EGDWPVPLEES---IKLEQEEER 1527

```

Search completed: July 11, 2002, 15:50:43
 Job time: 381 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 15:41:02 ; Search time 22.7 Seconds

(without alignments)
633,774 Million cell updates/sec

Title: US-09-687-230-2

Perfect score: 3073
Sequence: 1 MGKHKHKSDKHLVEEYK...PENMICKLGPSSKCLINK 589Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database : Issued Patents MA:*

- 1: /cgn2_6/prodata/2/1aa/5a_COMB.pep:*
- 2: /cgn2_6/prodata/2/1aa/5b_COMB.pep:*
- 3: /cgn2_6/prodata/2/1aa/5a_COMB.pep:*
- 4: /cgn2_6/prodata/2/1aa/5b_COMB.pep:*
- 5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
- 6: /cgn2_6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3073	100.0	589	4 US-08-942-008-2	Sequence 2, Appl1
2	198	6.4	2441	1 US-08-194-468-2	Sequence 2, Appl1
3	198	6.4	2441	3 US-08-961-739-2	Sequence 2, Appl1
4	190.5	6.2	2414	1 US-08-227-536-2	Sequence 2, Appl1
5	190.5	6.2	2414	5 PCT-US95-04682-2	Sequence 2, Appl1
6	181.5	5.9	1872	1 US-08-188-582-14	Sequence 14, Appl1
7	181.5	5.9	1872	1 US-08-646-715-14	Sequence 14, Appl1
8	181.5	5.9	1893	1 US-08-188-582-11	Sequence 11, Appl1
9	181.5	5.9	1893	1 US-08-646-715-11	Sequence 11, Appl1
10	150.5	4.9	1312	2 US-08-687-080-51	Sequence 51, Appl1
11	149.5	4.9	1312	2 US-08-592-126-148	Sequence 148, Appl1
12	147.5	4.8	1588	5 PCT-US93-07261-11	Sequence 11, Appl1
13	147.5	4.8	1588	5 PCT-US93-07261-16	Sequence 16, Appl1
14	144	4.7	1886	4 US-08-938-105-3	Sequence 3, Appl1
15	142	4.6	1898	2 US-08-056-200-94	Sequence 94, Appl1
16	142	4.6	1898	2 US-08-800-644-94	Sequence 94, Appl1
17	142	4.6	3248	1 US-08-353-700-1	Sequence 1, Appl1
18	142	4.6	3248	5 PCT-US95-16216-1	Sequence 1, Appl1
19	141.5	4.6	2482	1 US-08-328-254-6	Sequence 6, Appl1
20	140	4.6	1939	4 US-09-310-187A-1	Sequence 1, Appl1
21	138.5	4.5	765	2 US-08-663-112-2	Sequence 2, Appl1
22	133.5	4.3	1161	4 US-09-327-536-2	Sequence 2, Appl1
23	132.5	4.3	1388	2 US-08-685-576-4	Sequence 4, Appl1
24	130.5	4.2	1618	1 US-07-853-913-4	Sequence 4, Appl1
25	129.5	4.2	1829	4 US-09-157-420-1	Sequence 1, Appl1
26	128.5	4.2	816	2 US-08-533-306A-6	Sequence 6, Appl1
27	128.5	4.2	816	2 US-08-742-923A-6	Sequence 6, Appl1

28	128.5	4.2	885	2 US-08-533-306A-4	Sequence 4, Appl1
29	128.5	4.2	885	2 US-08-742-923A-4	Sequence 4, Appl1
30	128.5	4.2	1112	2 US-08-714-402-2	Sequence 2, Appl1
31	128	4.2	1388	4 US-09-572-191-2	Sequence 2, Appl1
32	127	4.1	724	4 US-09-307-143-6	Sequence 6, Appl1
33	126.5	4.1	976	4 US-09-104-324B-4	Sequence 4, Appl1
34	125	4.1	1507	3 US-08-929-329-5	Sequence 5, Appl1
35	124.5	4.1	1388	2 US-08-685-576-1	Sequence 1, Appl1
36	124	4.0	1128	4 US-08-923-992A-6	Sequence 6, Appl1
37	124	4.0	1346	1 US-08-471-033-23	Sequence 23, Appl1
38	124	4.0	1346	2 US-08-471-044-23	Sequence 23, Appl1
39	124	4.0	1346	2 US-08-463-483A-23	Sequence 23, Appl1
40	124	4.0	1346	2 US-08-471-046A-23	Sequence 23, Appl1
41	124	4.0	1346	2 US-08-470-566B-23	Sequence 23, Appl1
42	124	4.0	1346	2 US-08-469-334-23	Sequence 23, Appl1
43	124	4.0	1346	3 US-09-300-529-23	Sequence 23, Appl1
44	123.5	4.0	532	1 US-08-285-440-5	Sequence 5, Appl1
45	123.5	4.0	532	1 US-08-630-349-5	Sequence 5, Appl1

ALIGNMENTS

RESULT 1
US-08-942-008-2
Sequence 2, Application US/08942008
Patent No. 6133419
GENERAL INFORMATION:
APPLICANT: Braselmann, Sylvia
TITLE OF INVENTION: Nucleotide Sequences that Encode
Phosphatidylinositol 3' Kinase Associated Proteins and
TITLE OF INVENTION: Uses thereof
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESSES:
ADDRESSEE: ONYX Pharmaceuticals, Inc.
STREET: 3031 Research Drive
CITY: Richmond
STATE: CA
COUNTRY: USA
ZIP: 94806
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942,008
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Glotta, Gregory
REGISTRATION NUMBER: 32,028
REFERENCE/DOCKET NUMBER: ONYX1027
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 262-8710
TELEFAX: (510) 222-9758
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 589 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-942-008-2

Query Match 100.0%; Score 3073; DB 4; Length 589;
Best Local Similarity 100.0%; Pred No. 5, 1e-262;
Matches 589; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKHKHKSDKHLVEEYKPKLVKVGNGVTELTSGSSGSDSLFEDKNDHDKHND 60
Db 1 MGKHKHKSDKHLVEEYKPKLVKLVKVGNGVTELTSGSSGSDSLFEDKNDHDKHND 60

```

OY 61 RRRKKKGGKQIPGEEKGRKRRVKEDEKKRRDRVENEAEKDLQCHAPVRLDLPPEKP 120
DB 61 RRRKKKGGKQIPGEEKGRKRRVKEDEKKRRDRVENEAEKDLQCHAPVRLDLPPEKP 120
OY 121 LTSSIAKQEEVQOTPLQELNLMROLQKDPASAFSPYVDFIAPGSMITIKHMDST 180
DB 121 LTSSIAKQEEVQOTPLQELNLMROLQKDPASAFSPYVDFIAPGSMITIKHMDST 180
OY 181 MEKIKNNQYOSIEELKKNFKLMTNMIYKPEITYYKAAKLLHSGKILISOERISL 240
DB 181 MEKIKNNQYOSIEELKKNFKLMTNMIYKPEITYYKAAKLLHSGKILISOERISL 240
OY 241 KOSIDPMADLOKTRKOKDGTDSQSGEDGCGQOREREDSGDAEAAFKSPSKENKKDKD 300
DB 241 KOSIDPMADLOKTRKOKDGTDSQSGEDGCGQOREREDSGDAEAAFKSPSKENKKDKD 300
OY 301 MLEDFKSNLNEREQDLDRIYKESGKILTRLVNSQCEFEERKRDGTTGLLHPVPI 360
DB 301 MLEDFKSNLNEREQDLDRIYKESGKILTRLVNSQCEFEERKRDGTTGLLHPVPI 360
OY 361 VEEPGYCLVRLGTMTRLOSQVNTLOGFEKEDKRNKVTPLVLYNGPYSSAYPHYSTFAN 420
DB 361 VEEPGYCLVRLGTMTRLOSQVNTLOGFEKEDKRNKVTPLVLYNGPYSSAYPHYSTFAN 420
OY 421 ISKDDSDLIYSTYGEDSDLPSPDSIHEPLATCQDYPYVADSLDVLTKGHSRTLQEME 480
DB 421 ISKDDSDLIYSTYGEDSDLPSPDSIHEPLATCQDYPYVADSLDVLTKGHSRTLQEME 480
OY 481 MSLPDEGHTRTLDTGKEMEQITEVEPPGRDLSSTODRLIAKAVTNGVPEVDSSEA 540
DB 481 MSLPDEGHTRTLDTGKEMEQITEVEPPGRDLSSTODRLIAKAVTNGVPEVDSSEA 540
OY 541 EIFOKKLDETRRLRLQEAQNERLSTRPPGMICLGPSSSEKILINK 589
DB 541 EIFOKKLDETRRLRLQEAQNERLSTRPPGMICLGPSSSEKILINK 589

RESULT 2
US-08-194-468-2
; Sequence 2, Application US/08194468
; Patent No. 5750336
; GENERAL INFORMATION:
; APPLICANT: Montminy, Marc R.
; TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Preilly, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,468
; FILING DATE: 10-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9672
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)-546-4737
; TELEFAX: (619)-546-9392
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2441 amino acids

```

```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-194-468-2

Query Match          6.4%; Score 198; DB 1; Length 2441;
Best Local Similarity 19.9%; Pred. No. 4, 1e-08;
Matches 139; Conservative 93; Mismatches 219; Indels 248; Gaps 29;

OY 32 NEVFLSTGSSGSHSSLEDKNDKNDKRRKRRKKEKQIPGEEKGRKRRVKEDEKKRRDRVENEAEKDLQCHAPVRLDLPPEKP 88
DB 971 NKVPPTSTVTAEMTSS--QDGPDPVPMLEMTKEVQTDADPEEPTESSKGEPRSEMEEDLQ 1028
OY 89 ---KKRRDRVENEAEKDLQCHAPVRLDLPPEKPILTSIAKQEEVQOTPLQELNLMROLQKDPASAFSPYVDFIAPGSMITIKHMDST 134
DB 1029 GSSQVKEETDITTEQKSEP-----MEVEKRPDYVEVKEKEEENSSNTASQSTSPS 1079
OY 135 -----PLQELNLMROLQKDPASAFSPYVDFIAPGSMITIKHMDST 181
DB 1080 QPRKRTFAPPELRQALMPTLEALYKQDPESLPFPQVDPQLGIDPEDIYKNNPMDLSTI 1139
OY 182 KEKIKNNQYOSIEELKKNFKLMTNMIYKPEITYYKAAKLLHSGKILISOERISL 241
DB 1140 KRKLDTGOYQEWQVVDVRLMFNFMALYKRTSKVYKFCSKL---AEVEQER-IDPVM 1194
OY 242 QSIDPMADLOKTRKOKDGTDT-----SQSGEDGCG--QREERE 277
DB 1195 QSLGVCQG---RKTEFSPQTLCCYQKQLCTIPRDAAYSYQNNYHFGCKGFTLEQGENV 1250
OY 278 DSGDAEAAFKSPSKENKKDKMDLEDKFSNNLE-----RQEQDL----- 318
DB 1251 TLGD-----DPSQPTTISKQDEKK-KNDLTDPPEPVDCKEGRKMHQICVLHVDI 1302
OY 319 -----DRIVKESG-----GKILTRLVNSQCEFEERKRDGTTGLLHP 356
DB 1303 WPSGFCVDCNCLKTRGPRPKENKFSAKRLQTRNLGHLDRVANKRLRQN-----HP 1353
OY 357 VDPYIEBPGYCLVRLGTMTRLOSQVNTLOGFEKEDKRNKVTPLV---YLVNGPYSSAYPH 413
DB 1354 -----EAGEVFVRVVAAS-----DKTEVFKPMKMSRRVDSGEMSESPFY 1392
OY 414 -----YDSTANISKDSDLI-----YSTYGEDSDLPSPDSIHEPLAT 451
DB 1393 RTKALFAEEIDGVVCFEGHNVDTLALPHQIGCVIISYLD-----SHFRPR 1444
OY 452 C-----QDYPYVADSLDVLTKG-----HSRTLQEME 480
DB 1445 CLRTAVYHEILIGYLEYKVLVYTAHIMACPRSGDYYIFCHPPODKITPKRLQEMV 1504
OY 481 MSLPDEGHTRTLDTGKEMEQITEVEPPGRDLSSTODRLIAKAVTNGVPEVDSSEA 540
DB 1505 KKMIDKAFARLIINDYKDI-----FKQANEDRLTSAKELPYR-----EG 1543
OY 541 EIFOKKLDETRRLRLQEAQNERLSTRPPGMICLGPSSSEKILINK 589
DB 1544 DFWPVNLEES--IKLEQEEERKKEESTAASETPEGS 1579

RESULT 3
US-08-961-739-2
; Sequence 2, Application US/08961739A
; Patent No. 6063583
; GENERAL INFORMATION:
; APPLICANT: Montminy, Marc R.
; TITLE OF INVENTION: Methods for Treating Diabetes Mellitus
; FILE REFERENCE: SAK1650-1
; CURRENT APPLICATION NUMBER: US/08/961,739A
; CURRENT FILING DATE: 1997-10-31
; EARLIER APPLICATION NUMBER: US 194,468
; EARLIER FILING DATE: 1994-02-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0

```



```

0Y      499 MEQIEVEPPGG-----LD-----SSTODRLIAKAVN 527
Db      1444 DDYIHCHPDDQIKIRPKRLOEWYMKMLDKAVSERIVHDYKDIFKFOEDBRLTSAKELPY 1503
0Y      528 FGVPVPEFDESEAELEFKRIJDETETRLRLRELQANER 564
Db      1504 F-----EGDFWPNVLEEES---IKLELGEEBEER 1527

RESULT      5
PCT-US95-04682-2
; Sequence 2, Application PC/TUS9504682
; GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION
TITLE OF INVENTION: FACTOR P300 AND USES OF P300
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
STREET: Ten Post Office Square
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04682
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/227,536
FILING DATE: 14-April-1994
ATTORNEY/AGENT INFORMATION:
NAME: Holliday C. Helne, Ph.D.
REGISTRATION NUMBER: 34,346
REFERENCE/DOCKET NUMBER: DFCI-308XG999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-2290
TELEFAX: (617) 451-0313
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2414 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-04682-2

Query Match          6.2%; Score 190.5; DB 5; Length 2414;
Best Local Similarity 20.4%; Pred. 1.9e-07;
Matches 130; Conservative 83; Mismatches 249; Indels 175; Gaps 26;

0Y      38 SNGSGHDSLSFEJDNDNHKKHKRRKKRKKEKKOPGEEKGRKKRRRVYEDKKKKRRDRNV 97
Db      956 SYSSSTEVSNOAIAEK---OPSOEVRMEAAWEVDQPEPADTQPDISSESKEVEDCKMNSTET 1012
0Y      98 ENEAEKDJCAHAVRLDLPPEKPLTSLA----KOEVEQTPLEALNOLMROLORKDP5 153
Db      1013 E-EERSTDELTKIEEDDGSTATGSSPAAGSKKKIFKEPELRQLAMPTLEALYNODE 1071
0Y      154 AF-FSEPVTDTI--ARGISMIIKHPPDWSTKKEKIKNNDYOSIELKDNFKIMACTNAMTY 210
Db      1072 SLTFRPQVDPOLLGIPIDYDIYKSPMDISTIKRKIDTGQYQEPWQYVDDIWIMFNAMLY 1131
0Y      211 NKPEITYYAAAKKLHSKGKILISOEIGSLKSIDEMA-----DLQTRKQ 256
Db      1132 NKRITSVHYKYSGL---SEVEQE-IDPVWSLGICCGRKLFPSFPOLLCYGKQLCTIP 1186

```

```

QY 257 KGDITSGSGDGC-----WQERERSDDAAHAFKPSKEN-KKKDKMLEDKF--K 307
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1187 RDATYTSQNFYHFCSEKCFNEIGESVSLDDPSQFQTTINKFQPSKRANDTLDELVE 1246E
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 308 SNLNEREOEL-----DRYKESG-----GKLTFRLLY 334
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1247 CTBGRKMHQICVLHHEILMPAGFVCGCLKTSAPTRKENKFSARLPSRLGTLELNRV 1306E
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 335 NSQCEFERKRDGTTTGLHPVDPIYGEFGYCLVRLGHTTGRHSGVNTLOGFEEDKRN 394
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1307 NDFLRQNHPSGSEVTVRVYVHASDKTYE-----YKGMKARFVSG-EMASSF----- 1353S
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 395 KVTPLVLYNTGPRYSVAPHYDSFPAFNSKDSDLIY-----SYGSDSLPDS----- 442
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1354 -----PYTKA-----LFAFEIDGVCLCFPGMHQDETSGDCPPRQRRVYISY 1397E
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 443 -FSIHEPLATC--QDYRVYVADSLDLVTLKGHSRTLOEMENSLPEDGHTRLTDGKE 498
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1398 LDSVHFEPKCLTRVAVYHEILL-GYLEYVKKLGYT-----TGHVACPSPSG 1443S
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 499 MEQITVEPPEP-----LD-----SSTDDRIAAKATYN 527
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1444 DDYIFCHPQPKQIPKPKRLQEWYKKMLDAVSRIVHDYKDIQVATDRILTSAKELPY 1503S
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 528 FGVPVEPDSSEAEIFQKKIDETTRTLRLERQEAQNER 564
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1504 F-----EGDPFWNVLEES---IKELQDEEER 1527
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
US-08-188-582-14
: Sequence 14, Application US/08188582
: Patent No. 5534410
: GENERAL INFORMATION:
: APPLICANT: Tjian, Robert
: APPLICANT: Comali, Lucio
: APPLICANT: Dynalac, Brian D.
: APPLICANT: Hoey, Timothy
: APPLICANT: Ruppert, Siegfried
: APPLICANT: Tanese, Neoko
: APPLICANT: Wang, Edith
: APPLICANT: Weinzierl, Robert O.J.
: TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
: TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE
: NUMBER OF SEQUENCES: 36
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/188,582
: FILING DATE: 28-JAN-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Osman, Richard A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1872 amino acids
:

```


TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-188-582-14

Query Match 5.98; Score 181.5; DB 1; Length 1872;
Best Local Similarity 21.0%; Pred. No. 7.8e-07;
Matches 118; Conservative 81; Mismatches 195; Indels 167; Gaps 21;

QY 6 KKHSDKHLYEEYERKPLKLVKGVNEVTELTSGSSGHDSLFEDKNDKDKRRKK 65
DB 1354 KRRVGTTHVDYLNRRHKSIRRRTPMVTLSLESIIINDMDLPTTFPHVPAKV 1413
QY 66 RKKGEKOIPGEKGRKRRYKEDKKR---DRDR-----VENEKDLQCHAPVR-- 112
DB 1414 VKDYKII---TRPMDIQTLEENVRKRLYPSREFREHLELVKNSATYNGPKHSLQIS 1470
QY 113 ---LDLPEKPLTSSLAKEVEEQTPLQALNOLMROLQKRPASAF----- 155
DB 1471 QSMDLDCDEK-----LKEKEDKLARLEKAINPL---LDDDDQVAFSFIIDNIYTKMA 1521
QY 156 -----FSPYTDPLAGYSMTIKHPMDFSTMKIKKNDYQSIIEELKDNFKLCTNAMI 209
DB 1522 VPDSWPFHHPVNKKFVVDYKVIYNPMDLETIRKNISKHKYQSRRESFLDDVNLILANSVK 1581
QY 210 YNKPETIYYKAKKLLHSGMKILSQ--ERIOSLK-----QSIDFMADLQ 251
DB 1582 YNGPESQYTKAQEIIVNCYQTLTEYDEHLTQLEKDICTAKEALEAELESLEDPMPGPR 1641
QY 252 KTRKQKGTDTQS---GEDGCGWQREERDS--GDAEHAFFSKSPKSKKKDKD----- 300
DB 1442 YTPQPPPLDYDNTSLMSRDSVQDESNNMSVLDIPSATPEKQYTOREEDGDGLADEEE 1701
QY 301 -----MLEDKFKSNLNEREOOLDRIKESGKLTFRILVNSQCEFERRRKPDGTTT 350
DB 1702 GTVOQPAQSVLYEDLMS---EGEDDEBDAGSDEGDN---PFSAIQLSESGSDSDVG 1733
QY 351 LGLHPVDPIYGEPCYCLVRLGTTGRLOSGVNTLQGFEDKRRKVPVLY----- 401
DB 1754 SGGIRPQPRM-----LQE--NT---RMDMENESMSTYEGDGGASH 1791
QY 402 -----LNYGYSSYAPH---YDSTFANI----- 421
DB 1792 GLEDSNISTGSEYEPDPKSNTOPTSFSSIGGYEVSEEEDEEERSGPSVLSQVHLE 1851
QY 422 SKDSDLIYTYGHDSPD 442
DB 1852 DEEDSEDFHSTAG-DSDLOSD 1871

RESULT 7
US-08-646-715-14
Sequence 14, Application US/08646715
Patent No. 563786
GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comal, Lucio
APPLICANT: Dymact, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California

COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,715
FILING DATE: 09-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/188,582
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1872 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-646-715-14

Query Match 5.98; Score 181.5; DB 1; Length 1872;
Best Local Similarity 21.0%; Pred. No. 7.8e-07;
Matches 118; Conservative 81; Mismatches 195; Indels 167; Gaps 21;

QY 6 KKHSDKHLYEEYERKPLKLVKGVNEVTELTSGSSGHDSLFEDKNDKDKRRKK 65
DB 1354 KRRVGTTHVDYLNRRHKSIRRRTPMVTLSLESIIINDMDLPTTFPHVPAKV 1413
QY 66 RKKGEKOIPGEKGRKRRYKEDKKR---DRDR-----VENEKDLQCHAPVR-- 112
DB 1414 VKDYKII---TRPMDIQTLEENVRKRLYPSREFREHLELVKNSATYNGPKHSLQIS 1470
QY 113 ---LDLPEKPLTSSLAKEVEEQTPLQALNOLMROLQKRPASAF----- 155
DB 1471 QSMDLDCDEK-----LKEKEDKLARLEKAINPL---LDDDDQVAFSFIIDNIYTKMA 1521
QY 156 -----FSPYTDPLAGYSMTIKHPMDFSTMKIKKNDYQSIIEELKDNFKLCTNAMI 209
DB 1522 VPDSWPFHHPVNKKFVVDYKVIYNPMDLETIRKNISKHKYQSRRESFLDDVNLILANSVK 1581
QY 210 YNKPETIYYKAKKLLHSGMKILSQ--ERIOSLK-----QSIDFMADLQ 251
DB 1582 YNGPESQYTKAQEIIVNCYQTLTEYDEHLTQLEKDICTAKEALEAELESLEDPMPGPR 1641
QY 252 KTRKQKGTDTQS---GEDGCGWQREERDS--GDAEHAFFSKSPKSKKKDKD----- 300
DB 1442 YTPQPPPLDYDNTSLMSRDSVQDESNNMSVLDIPSATPEKQYTOREEDGDGLADEEE 1701
QY 301 -----MLEDKFKSNLNEREOOLDRIKESGKLTFRILVNSQCEFERRRKPDGTTT 350
DB 1702 GTVOQPAQSVLYEDLMS---EGEDDEBDAGSDEGDN---PFSAIQLSESGSDSDVG 1733
QY 351 LGLHPVDPIYGEPCYCLVRLGTTGRLOSGVNTLQGFEDKRRKVPVLY----- 401
DB 1754 SGGIRPQPRM-----LQE--NT---RMDMENESMSTYEGDGGASH 1791
QY 402 -----LNYGYSSYAPH---YDSTFANI----- 421
DB 1792 GLEDSNISTGSEYEPDPKSNTOPTSFSSIGGYEVSEEEDEEERSGPSVLSQVHLE 1851
QY 422 SKDSDLIYTYGHDSPD 442

DB 1852 DEDESEDFHSIAG-DSDDLSD 1871

RESULT 8

US-08-188-582-11

Sequence 11, Application US/08188582

Patent No. 5634410

GENERAL INFORMATION:

APPLICANT: Tjian, Robert

APPLICANT: Comal, Lucio

APPLICANT: Dylact, Brian D.

APPLICANT: Hoev, Timothy

APPLICANT: Rupert, Siegfried

APPLICANT: Tanese, Naoko

APPLICANT: Wang, Edith

APPLICANT: Weinzierl, Robert O.J.

TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS

TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE

NUMBER OF SEQUENCES: 36

CORRESPONDENCE ADDRESS:

ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/188,582

FILING DATE: 28-JAN-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Osman, Richard A

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: A-57650-2/AUT/RAO

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 1893 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-188-582-11

Query Match 5.9%; Score 181.5; DB 1; Length 1893;

Best Local Similarity 21.0%; Pred. No. 8e-07;

Matches 118; Conservative 81; Mismatches 195; Indels 167; Gaps 21;

DB 6 KKHSDKHLVEEYKPLVLKVGNEVTELTSGSSGSDSLFEDKNDHKHDKRRKK 65

DB 1375 KRRVGTWVHCYDILNRPKSHRRRTPDPMVTLTSLIINDMRDLPNTPTPVNAKY 1434

DB 66 RKKGGEQIGEGEKKRRRRVKKEDKKR---DADR-----VENAEKDLQCHAVR-- 112

DB 1435 VADYVITL---TRPMDLQTLREVRKRLTPSRREPREHLELIVKNSATYNGPKHSLTQLS 1491

DB 113 ---LDLPEKPLTSLAKOEVEQPLQALNOLMOLQKRDPSAF----- 1551

DB 1492 QSMULDLCDEK-----LKRKEKDLAKLEKAINPL---LDODDOVAFSFLDNIYQKMA 1542

DB 156 -----FSPVYDFIAPAGYSMTIKHPMDFSTMEKIKNNDYOSIEELKDNFKIMCTNAMI 209

DB 1543 VDSWPFHHPVKKFVDDYKYIVNPMDELTTRKNISKHKYQSRESFLDDVNLILANSVK 1602

DB 210 YNKPETIYKAANKLIHSGMKILSQ--ERIOSLK-----QSIDPMADIQ 251

DB 1603 YNPGESQYTKTAQETIYVWCYQTLREYDEHLTQLEKDTCKAKPALEAELESLDPMTPGP 1662

DB 252 KTRKQKDGTDTSOS---GDDGCGOWEREDS--GDAEAHAFSPSKENKKKKD----- 300

DB 1663 YTPQPPDLTYDTNLSMSKDAVAFODESMNSVLDIPSAIPPEKQVTOEGEDGGLADEEE 1722

DB 301 -----MLEDKFSNNLEREOELDRIVKESGKLTRELVNSOCEFFERRKPDGTTT 350

DB 1723 GTVQOPQASVLEDDILMS---EGEDEEDAGSDEEDN-----PESAIQLSSGSDSDVG 1774

DB 351 LGLHPVDPIVGEPCYCLVRLGNTTGRLOSGVNTLQGFEDKRNKVTPLY----- 401

DB 1775 SGGIRKQPRM-----LOE--NT-----RMDMENESMNSYEGDGEASH 1812

DB 402 -----LNYGPSYAPH---YDSTPANI----- 421

DB 1813 GLEDSNISTGSTEEDPPKNTQDTSPSSIGYEVSEEDDEEDRQSGPSVLSOVHUSE 1872

DB 422 SKDSDSLIYSTGEDSLPSD 442

DB 1873 DEDESEDFHSIAG-DSDDLSD 1892

RESULT 9

US-08-646-715-11

Sequence 11, Application US/08646715

Patent No. 5637686

GENERAL INFORMATION:

APPLICANT: Tjian, Robert

APPLICANT: Comal, Lucio

APPLICANT: Dylact, Brian D.

APPLICANT: Hoev, Timothy

APPLICANT: Rupert, Siegfried

APPLICANT: Tanese, Naoko

APPLICANT: Wang, Edith

APPLICANT: Weinzierl, Robert O.J.

TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS

TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE

NUMBER OF SEQUENCES: 36

CORRESPONDENCE ADDRESS:

ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/646,715

FILING DATE: 09-MAY-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/188,582

FILING DATE: 28-JAN-1994

ATTORNEY/AGENT INFORMATION:

NAME: Osman, Richard A

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: A-57650-2/AUT/RAO

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 1893 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-646-715-11

Query Match 5.9%; Score 181.5; DB 1; Length 1893;
Best Local Similarity 21.0%; Pred. No. 8e-07;
Matches 118; Conservative 81; Mismatches 195; Indels 167; Gaps 21;

QY 6 KKHSDHLYEEVEKPLKLVKGVNEVTELGSSGHSLSLEFDKNDHKRRKK 65
DB 1375 KKRHVGTTHCDYLNRRHKSIIHRRTPMVLSSILESIINDMDLNTPEFHPVAKV 1434
QY 66 RKKEKQIPGEKRRRRKRVKEDKKR--DRDR-----VENEKDLQCAPR-- 112
DB 1435 VKDYKTI--TRPMDLQTLNENVRKRLYPSREPREHLELYNSATYNGPKHSITQIS 1491
QY 113 ---LDLPEKPLTSLAKQEEVEQTPQOALNOLMROLQRKDPGAF----- 155
DB 1492 QSMIDLQDEK-----LKEREDKLARLEKAIINPL---LDDDDQVAFSFIIDNIVTQKMA 1542
QY 156 -----FSPVTFDIADGYSMITKHPMDFSTMKEKIKNDYOSIEELKDNFKIMCTNAMI 209
DB 1543 VPDSMPFHHPVKKFVPDYKVIYNPMDLETIRKNISKHKYQSSRESFLDDVNIILANSVK 1602
QY 210 YNKPETIYKAAKKLLHSGCKIISQ--ERIQSLK-----QSIDPMADLO 251
DB 1603 YNGESQYTTAIDEIVAVCYQTLTEYDEHLLTOLEKICTAKALAEALLESIDPMTPGP 1662
QY 252 KTRKQKDGDTDSQ---GEDGGCWQREREDS--GDAEVAHAFKSPSKKKKKD----- 300
DB 1663 YTPQRPDLVYTNISLSMRASVQDSSNMVSDIPATPEKQVOTQCEGDDGLADEEE 1722
QY 301 -----MLEDKFSNNLEREQDLRIYVESGKTLRLVNSQCEFERRRKPDGTTT 350
DB 1723 GYVQOPQASVLYEDELMS---EGEDEDEADGSDGEGDN---PFSAIQLSESGSDGVG 1774
QY 351 LGLHPDPIYVGPBGYCLVRLGWTGRLOSGCVNTLOGFKEKDKNKVTPVLY----- 401
DB 1775 SGGRPQOPM-----LQE--NT---RDMENESMSTYEGDGEASH 1812
QY 402 -----LNGPYSSYAH--YDSTFANI----- 421
DB 1813 GLEDSNYSYSYEPDPKSNMTPQTSFSSIGYEVSEEEDEEBOQSGPSVLSQVHLSE 1872
QY 422 SKDSDSLVSTYGEDSLPSD 442
DB 1873 DEEDSEDFHSIAG-DSOLDSD 1892

RESULT 10

US-08-687-080-51
Sequence 51, Application US/08687080
Patent No. 5965427

GENERAL INFORMATION:

APPLICANT: Gregory Dolganov

TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof

NUMBER OF SEQUENCES: 175

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dehlinger & Associates

STREET: 350 Cambridge Avenue, Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/687, 080

FILING DATE: 17-JUL-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,126
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 1312 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: TRANS. OF RAD50 cDNA (SEQ. 54), NT.
US-08-687-080-51

Query Match 4.9%; Score 150.5; DB 2; Length 1312;
Best Local Similarity 17.8%; Pred. No. 0.00024;
Matches 133; Conservative 124; Mismatches 260; Indels 229; Gaps 29;

QY 2 GKHKHKKSDKILYEEVEKPLKLVKGVNEVTELGSSGHSLSLEFDKNDHKRRKK 43
DB 200 GKVKVEYQWELKYLKQKACEIRDOITSKEAOLTSSEKIYKVENEDLPKLNRLKEI 259
QY 44 HDSLEPEDKNDHDKRRKKRKKRKEKQIPGEKGR-----RRVKE 87
DB 260 HNSKTKMLKNDNEKALDSKKQKMEKDNSEL--EKKKVEFGQTDQGLNDLYNNHQPATVE 317
QY 88 DKRR-----RDRRVENEA-----EKDLQCHAPVRLDLPPEK-----PLTSLAKO 128
DB 318 KERKLVDCRHELEKLNKESRLNLOEKSELLVGRLOQADRHQEHIRARDSLQSLAQ 377
QY 129 EYV---EQTPLEA---LNQMLROLQRKDPGAFSFPVTDIAGYSMITKHPMDFSTM 181
DB 378 LEIDGFERGPFSEKQIKNFHKLVREKQ-----GEAKTANQIANDFAE 420
QY 182 KERKNNQOSIEELKDNF-KIMCTNAMIYNKPE---TIYK-----AAKKLHSGK 230
DB 421 KETLKQKQIDETRDKKTGLRIIELEKSELISKQNELKAVKTELOQLBSSDRILELDOE 480
QY 231 ILSQER-----ISLQKQSI---DPMADLQTRKQKDGDTDSQSGEDGGCWQR- 275
DB 481 LKKAERELSKAEKSNVVELKMEVISLQNEKADLRLTLKLDQEMEQLNHHHTTRQEM 540
QY 276 -HEDSGDAHAHFKSPSK------NKKDKDMLKFK------ 307
DB 541 LTRKADKDEQIRKIKSRISDELTLGLYFPNKKQLEMLHASKREINQTRDLKLNKE 600
QY 308 -----SNLREQEQL---DRIYKESGKTLFRVLNSQCEFERRRKPD----- 346
DB 601 LASSQKNKHINNELKRRKEQLSYEDKLFVDCGQ-----DFSSDLRLKEELEKSSK 654
QY 347 -----GTTTGLLHPVDPIYGEPGYC---LVRLGNTGRLOSQVNTLOG----- 387
DB 655 QRAMLGATAV-YSQFITQITDENOSCCPVQCFQTEALEQEVISDLSQSLRLAPDKL 713
QY 388 -----FKEDRNK---VTPVLYNIGYSSYAHHYSTFANISKDSDSLVSTYGEDS 437
DB 714 STSELKKEKRRDEMLGLVPRKOSTIIDKEKEIPELRLKLNQVNRDQRLKNDIEQET 773
QY 438 DLPDSFISHEFLATQDYPYVNASDLVLTGKGSRTLOEMEMSLPEDEGHTTLDTGK 497
DB 774 LGGTIMPEESAKVC-----LTDVTLMERQWELKQVIERKIAQQAQKQIDIDLR 823

MOLECULE TYPE: protein
; PCT-US93-07261-11

Query Match	4.8%;	Score 147.5;	DB 5;	Length 1588;
Best Local Similarity	19.5%;	Pred. No. 0.0006;		
Matches 131;	Conservative 109;	Mismatches 238;	Indels 193;	Gaps 32

```

0Y 50 EDDDDDKHND--RRKKKKKGGKQIPGEGKR-----KRRVYEDKKRRDR 96
Db 29 QKNDQKAKADLTKKSSQDSSSEKSLKEKVNBEALKEKENETLKKLENOKEKEKKA 88
0Y 97 VENEAEKEDLOCHAPVLD--LPPEKPLTSSLAKEOEVEBTPLOELNQLMROLOKKDPSA 154
Db 89 IKONNDEALKNNKNGNDKKDKKIVPKKP-----ESVEKDLKEMELKEK-- 129
0Y 155 FFSFVTDFTADPQYSMIITHPMDFSTMKKEIKNNDYOSI--EELDNKIKCTNMIATNK 212
Db 130 -----EFIKHLDKYEERKEKRRNWILRSIRNRKLRKEIQLKELNANOLESA 175
0Y 213 PETIYYKAKK-----LHSGMKILSOERIKOSIDIPMDLOKTRKK-----DG--- 259
Db 176 IMELERRASRRPMVVKMORGMK---DEVDWIKIYKDEQAEKNGTKDEEIKDKGDYEE 232
0Y 260 -TDTSSQSG-----EODGCGNOREDSGDMAFAAFSP----- 290
Db 233 IYETNFKYGRNALGELDEYERIEYKRRYLLKEDGDLKDYEEKLEETGYGFRKPEPTT 292
0Y 291 ----SKENKKDKMLIEDKFKSNLREBOQLDRI-VKESGKLTIRLVNOCFF--ERR 343
Db 293 RILVRRKRNKEOKKLEKDEKKEKLLIAEERDDEKKIKLSDSDKYVVPVNNKKSSFPDKR 352
0Y 344 KPDGTTT-----LGLHPDPIVGEYGCYLVRLGHTTGKLOSGVN-----TLGFKEDKR 393
Db 353 APDKRTEFYRLSELPLPYPRKDN-----BLAVCGDSMDKVGKLIKLSFNPFKR--RR 405
0Y 394 NKVTEPLYLNTGPRYSYAPHYDSTFANISKD-----SDLI----- 429
Db 406 NKLKRRKMOELHKFKNNKKYOKLLERKRRNPGRPLNPIETIVIRPSLDMDKGENKSA 465
0Y 430 -----YTYG-----EDSDLPSPDSI--HEPLATQDYP--YVMADSLD-----VLT 468
Db 466 GHPEYOPTKGLKEYEESHVSKDQOLEHEPPTKPLPEYKGVHSREYQOLDEHPPTKLPYE 525
0Y 469 KGGHSRTLO---EMEMSLPE--DEGHTR--TLD-----TGKEMQITEV-----EPP 508
Db 526 KGVHSREYQOLNEVRODELPEYKGVHSREYQOLDEGSPSTKEYDQ--TELAKGKIDITNPH 584
0Y 509 GRDSDSTODRIALKAIVNFGVPVEFDSSEAEIIFORKLDETTPLRLAELO-----AON 582
Db 585 ESVDYDQTELAKKGDITN--KPHESVDYDOSELAKGKIDITNPHESVDYDQTELAKG 642
0Y 563 ERLSTRPPGMN 573
Db 643 KEVNTKPHENL 653

RESULT 13
PCT-US93-07261-16
Sequence 16, Application PC/TUS9307261
GENERAL INFORMATION:
TITLE OF INVENTION: PFEMP3 MALARIA ANTIGEN ANTIBODIES AND USES THEREOF
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: John H. C. Blasdale
STREET: One Glralda Farms
CITY: Madison
STATE: New Jersey
COUNTRY: USA
ZIP: 07940-1000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh

```

```

1      OPERATING SYSTEM: Macintosh 6.0.5
2      SOFTWARE: Microsoft Word 5.1a
3      CURRENT APPLICATION DATA:
4      APPLICATION NUMBER: PCT/US93/072661
5      FILING DATE: 19930805
6      PRIOR APPLICATION DATA:
7      APPLICATION NUMBER: US 07/927,531
8      FILING DATE: 07-AUG-1992
9      ATTORNEY/AGENT INFORMATION:
10     NAME: Blasdale, John H. C.
11     REGISTRATION NUMBER: 31,895
12     REFERENCE/DOCKET NUMBER: DX0288K
13     TELECOMMUNICATION INFORMATION:
14     TELEPHONE: 201-822-7398
15     TELEFAX: 201-822-7039
16     INFORMATION FOR SEQ ID NO: 16:
17     SEQUENCE CHARACTERISTICS:
18     LENGTH: 1663 amino acids
19     TYPE: amino acid
20     TOPOLOGY: linear
21     MOLECULE TYPE: peptide
22     ORIGINAL SOURCE:
23     ORGANISM: Plasmodium falciparum
24     STRAIN: Malayan Camp
25     PCT-US93-072661-16

```

Query Match	4.88;	Score 147.5;	DB 5;	Length 1663;
Best Local Similarity	19.58;	Pred. No. 0.00064;		
Matches 131;	Conservative 109;	Mismatches 238;	Indels 193;	Gaps 32

[illegible]

Db 585 ESVDYDQTELAKGKDTN--KPHESVDYDQSELAKGKDTNKPHEVDYDQTELAKG 642
QY 563 ERLSTRPPGN 573
: : : : :
Db 643 KEVTKKPHENL 653

RESULT 14
US-08-938-105-3
: Sequence 3, Application US/08938105
: Patent No. 6353151

: GENERAL INFORMATION:
: APPLICANT: Leinwand, Leslie A.
: APPLICANT: Vikstrom, Karen L.
: TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sheridan Ross P.C.
: STREET: 1700 Lincoln St., Suite 3500
: CITY: Denver
: STATE: CO
: COUNTRY: U.S.A.
: ZIP: 80203

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/938,105
: FILING DATE:
: CLASSIFICATION:

: ATTORNEY/AGENT INFORMATION:
: NAME: Crook, Manneil M.
: REGISTRATION NUMBER: 31,071
: REFERENCE/DOCKET NUMBER: 3595-4
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303) 863-9700
: TELEFAX: (303) 863-0223
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1886 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-938-105-3

Query Match 4.7%; Score 144; DB 4; Length 1886;
Best Local Similarity 19.5%; Pred. No. 0.0016; Indels 162; Gaps 27;
Matches 119; Conservative 111; Mismatches 217;

QY 47 SLEFDKNDHD-----KHKDRKKKKKKGEKQIDGEEK-----GR 80
: : : : :
Db 828 SLQEKNDLQVQADQEDNADAEKRDQIKKKQLKAVKMTRELEDEEMNMLTA 887
QY 81 KRRRVKED--KKRRDRVE--NEAEDLQCHAVRDLPEKFLTSLAKOEY----- 131
: : : : :
Db 888 KKRLEDECESELKKDIDDLTLTAKYKEKHATEN---KVKMLTEEMAGLDELITAKLK 943
QY 132 EQPLQALQMLQKQALQKPSAFSPVDTFAPGSMILIKHPDFTM--KEKIKNDY 190
: : : : :
Db 944 EKKALQEAHQADDDLOAEDEKY-----NTLTGSKVLE-- 977
QY 191 QSIEELDNFKMCTNMIYKPEITYYKAKKILSHGKILSOERISLQKSIDFMADL 250
: : : : :
Db 978 QVVDLGLSGLE-----QEKVYRMDLERAKRKLEGLD--LVOESIMDLND--KIQ 1025
QY 251 QKTRKQKQDGTDSQSGEDGCGWQREKRDSDGAENAHAKSPSKKKKKDKMLL----- 303
: : : : :
Db 1026 EERLKKE--FDISO-----QNSKLEEDQALQLOKTKLEKQAAIEELBELEERT 1076
QY 304 -----DKFSKNLREDEQDLRIYKESGKLTTRL---VNSOCFFERRKPD-----G 347

Db 1077 ARAKVEKLR--DLTRELEISERLEAGATSVQIEMNKKRAEFQKMRDRDEEATLQHE 1135
QY 348 TTLGLLHPDPIYVGPYCVLGMTTGRJQSGVTLQFKEDKKNKTPVILYINYPY 407
: : : : :
Db 1136 ATAALKRKHADSVAEIG-----EIDNLDVRKO--KLEKSKSEKLELDY 1180
QY 408 SSYAPHYDSTFANISKDDSLIYTYGEDSDLPSPDSIHEFLATCODYPYVMA----- 460
: : : : :
Db 1181 TSHMEQITKAKNLEKSRLL-----EDQANEYRVKLEAQSLNDFTTORAKIQTENG 1234
QY 461 -----DSLDDYITKG--GHSRTIQEMEMSLPDEGHTR-----TLDGKEMQDYE 504
: : : : :
Db 1235 ELAROLEEKALIMQLRGKLSYQOEDLKRL--EESGAKAALAHALQSAHDCDLR 1293
QY 505 VEPGRJDSSTODRLILKA---VTNGVPE---VFDSEAEIIFOKLDETRRLREIQ 558
: : : : :
Db 1294 EYEEEMAKAEIQRVLSKANSEYAVQWRTYEDAIQRTLELEAKKILAQRLQDAEAV 1353
QY 559 EAQNERLST 567
Db 1354 EAVNAKCS 1362

RESULT 15
US-08-056-200-94
: Sequence 94, Application US/08056200
: Patent No. 5616500

: GENERAL INFORMATION:
: APPLICANT: Steinert, Peter M.
: APPLICANT: Lee, Seung-Chul
: APPLICANT: Kim, In-Gyu
: APPLICANT: Chung, Soo-Il
: TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
: TITLE OF INVENTION: Methods of Using Same
: NUMBER OF SEQUENCES: 117
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe, Martens, Olson & Bear
: STREET: 620 Newport Center Drive, Sixteenth floor
: CITY: Newport Beach
: STATE: CA
: COUNTRY: U.S.A.
: ZIP: 92660

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/056,200
: FILING DATE: 30-APR-1993
: CLASSIFICATION: 435

: ATTORNEY/AGENT INFORMATION:
: NAME: Fedtlick, Michael F.
: REGISTRATION NUMBER: 36,799
: REFERENCE/DOCKET NUMBER: NIH054,001A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (714) 760-9502
: TELEFAX: (714) 760-9502
: INFORMATION FOR SEQ ID NO: 94:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1898 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-056-200-94

Query Match 4.6%; Score 142; DB 1; Length 1898;
Best Local Similarity 18.2%; Pred. No. 0.0024; Indels 100; Gaps 11;
Matches 64; Conservative 69; Mismatches 118;

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 15:32:31 ; Search time 55.8 Seconds

(without alignments)
1172.446 Million cell updates/sec

Title: US-09-687-230-2

Perfect score: 3073
Sequence: 1 MCKRHKRKHSDKHXYEVE.....PENNMLCLGPSSKCLINK 589

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3073	100.0	589	19	AAW37947
2	2995	97.5	667	22	AAO13504
3	2972.5	96.7	718	22	AAU1626
4	1710.5	55.7	405	22	AAW3712
5	1413.5	46.0	351	22	AAW3765
6	920.5	30.0	597	22	AAW3835
7	920.5	30.0	605	22	AAW40621
8	798.5	26.0	199	22	AAU16213
9	699.5	22.8	861	22	ABW3028
10	684	22.3	501	22	AAW3421
11	476	13.5	280	21	AAW41780

12	391	12.7	233	22	AAW5681	Human protein sequ
13	258.5	8.4	1058	22	AAW39231	Human polypeptide
14	258.5	8.4	1109	22	AAW41017	Human polypeptide
15	258.5	8.4	1109	22	AAW41018	Human polypeptide
16	258.5	8.4	1189	22	AAW39232	Human polypeptide
17	252.5	8.2	414	22	AAW63767	Human prostate can
18	252.5	8.2	715	22	ABG12634	Human prostate can
19	235.5	7.7	707	22	ABG21300	Novel human diagno
20	235.5	7.7	1173	22	ABG21301	Novel human diagno
21	229.5	7.5	2543	22	ABG21295	Novel human diagno
22	221	7.2	616	21	ABG28052	Novel human diagno
23	218	7.1	1430	22	ABW58602	Arabidopsis thalia
24	215	7.0	221	21	AAW43501	Human cancer assoc
25	212	6.9	245	22	ABW3828	Human cancer assoc
26	210.5	6.8	242	22	ABW44555	Human wound healin
27	208	6.8	2065	22	ABW44556	Human wound healin
28	208	6.8	2065	22	ABW44556	Human wound healin
29	198	6.4	2441	16	AAW70054	Human wound healin
30	198	6.4	2441	19	AAW40058	CREB binding prote
31	198	6.4	2441	22	AAW44252	Cellular transcrip
32	192.5	6.3	1649	21	AAW44555	Mouse nuclear CREB
33	191	6.2	1650	21	AAW27557	Human wound healin
34	191	6.2	1678	21	AAW27554	Human tumour suppr
35	190.5	6.2	2414	16	AAW84882	Human tumour suppr
36	190.5	6.2	2414	19	AAW40057	Human tumour suppr
37	189.5	6.2	1679	21	AAW27552	Transcription fact
38	188.5	6.1	1646	21	AAW27553	Human tumour suppr
39	188.5	6.1	1681	21	AAW27558	Human tumour suppr
40	187	6.1	1647	21	AAW27549	Human tumour suppr
41	187	6.1	1682	21	AAW27556	Human tumour suppr
42	187	6.1	1924	22	ABG06008	Human tumour suppr
43	186.5	6.1	1924	22	ABG06008	Novel human diagno
44	183	6.0	580	22	AAU16207	Novel human diagno
45	183	6.0	2234	22	ABG00985	Human novel secret

ALIGNMENTS

RESULT 1	AAW37947	AAW37947	standard: Protein: 589 AA.
ID	XX	XX	AAW37947;
AC	XX	XX	14-SEP-1998 (first entry)
DE	XX	XX	Phosphatidylinositol-3' kinase associated protein.
KW	XX	XX	Phosphatidylinositol-3' kinase associated protein; PI3K; PIKAP;
KW	XX	XX	human; signal transduction; cell growth; cancer; restenosis;
KW	XX	XX	therapy; diagnosis.
OS	XX	XX	Homo sapiens.
FT	XX	XX	Key
FT	XX	XX	Domain
FT	XX	XX	Region
FT	XX	XX	Location/Qualifiers
FT	XX	XX	151..313
FT	XX	XX	/note="bromodomain"
FT	XX	XX	516..589
FT	XX	XX	/note="PI3K p85 binding region"
PN	XX	XX	W09820126-A1.
PD	XX	XX	14-MAY-1998.
PD	XX	XX	01-OCT-1997;
PF	XX	XX	97WO-US15845.
PR	XX	XX	01-NOV-1996;
PR	XX	XX	96US-0030103.
PA	XX	XX	(ONVX-) ONVX PHARM INC.
PI	XX	XX	Brasellmann S;

DR WPI, 1998-286942/25.
DR N-PSDB; AAV29267.

XX New isolated phosphatidylinositol-3/kinase associated protein -
PT used to develop products for diagnosis and treatment of cell growth
PT disorders such as restenosis or cancer

PS Claim 10; Page 40-41; 52pp; English.

XX This polypeptide comprises human phosphatidylinositol-3' kinase
CC (PI3K) associated protein (PI3AP), a protein that binds to the
CC intermediate SH2 domain on the p85 regulatory subunit of PI3K, and
CC which exhibits a bromodomain. Its amino acid sequence was deduced
CC from a cDNA clone (see AAV29267) obtained from an HeLa library using
CC a yeast two-hybrid assay with PI3K p85 as bait. The invention
CC provides vectors containing nucleic acid sequences that encode
CC PI3AP or its fragments, host cells, methods for the expression of
CC PI3AP, and methods for using the products for the diagnosis and
CC treatment of cell growth disorders such as restenosis or cancer.
CC Also described is an assay for identifying agonists and antagonists
CC of PI3K regulation. These include mutant PI3APs that compete with
CC native PI3APs for binding to PI3K, antibodies, and nucleotide
CC sequences that can be used to inhibit or enhance PI3AP gene
CC expression. Transgenic and knock-out animals are also described.

SO Sequence 589 AA;

Query Match 100.0%; Score 3073; DB 19; Length 589;
Best Local Similarity 100.0%; Pred. No. 6.7e-248;
Matches 589; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 MGKRRKRRKSKHLVEEYVKKPLKLVKVGNEVTELTSGSSGHSSLFEDKNDHKDKKD 60
DB 1 mgkrrkrrkskdhlyeevkkplklvkvgnvteletsgssghsslfedkndhdkkd 60
OY 61 RRRRRKRRKGEKQIPGEEKRRRRVKKEDKKRRDRVNEAEKDLQCHAPYRLDIPPEK 120
DB 61 rrrrrkrrkgeqkqipgeekrrrrvkkedkrrdrvneaeekdlqchapyrldippe 120
OY 121 ITSSIAKQEEVQNPLOALNQLMRQKRPDPSPFVDFIAPGSMIIKHPMDFST 180
DB 121 itssiakeevqnploalnqlmrqkrdpdpfpvdfiapgsmiikhpmdfst 180
OY 121 ltslakveevqnploalnqlmrqkrdpdpfpvdfiapgsmiikhpmdfst 180
DB 121 ltslakveevqnploalnqlmrqkrdpdpfpvdfiapgsmiikhpmdfst 180
OY 181 MKEKIKNDYOSIEELKDNFKLMCTNAMIYKRPETIYKAAKKILHSGMKILISOER 240
DB 181 mkekikndyosieelkdnfkmlctnamiykrpetiyykaakkilhsgmkilisoer 240
OY 241 KQSIDFMA DLQTRKQKQGTDTOSGSEGGCQWQREERDSGDAEAFKSPSKENK 300
DB 241 kqidfmadlqtrkqkqgtdtotsgeggcqwqreerdsdgaefafkspskenk 300
OY 301 MEDFKSNLREDEQDLRIYKESGKLTIRLVNSOCEFERRRKPDGTTTGLLHPVPI 360
DB 301 mledfksnlredegdlriykesgkltirrvnsocfeferrrkpdgtttgllhpvp 360
OY 361 VGEPEYCLVRLQMTTGRQLQSGVNTLQGRKEDKRNKVTPLYLNTGPPYSYAHNT 420
DB 361 vgepeyclvrlqmttgrqlqsgvntlqgrkedrknkvtplylntgppysyahnt 420
OY 421 ISKSDSLIYTYGSDSLPDSFSIHEFLATQDYPYVADSLDLYLTGSGSRTLQEME 480
DB 421 isksdslitytygsdslpdsfsiheflatqodypyvadslldlyltgsgsrtlqem 480
OY 481 MSLEPDEGHTRTLDTGKNEQITVEVPPGRLLDSTODRLIALKAVTNGCVPEVDE 540
DB 481 mslepdeghtrtldtgkneqitvevppgrlldstodrlialkavtngcvpevde 540
OY 541 ELTFQKRLDETTLRLRLQEAQNRSLSTRPGNWCILGSSSRKCTILNK 589
DB 541 elftqkrldeptrllrlrlqeaqnrslstrpgnwcilgsssrkctilnk 589

```

RESULT 2

ID AA013504 standard; protein; 667 AA.

XX AA013504;

XX 06-NOV-2001 (first entry)

XX Human polypeptide SEQ ID NO 27396.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

XX WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001MO-US04927.

XX 26-FEB-2000; 2000US-0515126.

XX 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI, 2001-514838/56.

XX N-PSDB; AA193435.

XX Isolated nucleic acids and polypeptides, useful for preventing

XX diagnosing and treating e.g. leukaemia, inflammation and immune

XX disorders -

XX Claim 20; SEQ ID NO 27396; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_poc_sequences.

XX Sequence 667 AA;

Query Match 97.5%; Score 2995; DB 22; Length 667;
Best Local Similarity 99.5%; Pred. No. 2.6e-241;
Matches 575; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

OY 3 KKKKKKSDKHLVEEYVKKPLKLVKVGNEVTELTSGSSGHSSLFEDKNDHKDKKD 62
DB 18 kkkkkksdkhlyeevkkplklvkvgnvteletsgssghsslfedkndhdkkd 77
OY 63 RRRRRKRRKGEKQIPGEEKRRRRVKKEDKKRRDRVNEAEKDLQCHAPYRLDIPPE 122
DB 78 rrrrrkrrkgeqkqipgeekrrrrvkkedkrrdrvneaeekdlqchapyrldippe 137
OY 123 SSLAKQEEVQNPLOALNQLMRQKRPDPSPFVDFIAPGSMIIKHPMDFSTMK 182
DB 138 sslakveevqnploalnqlmrqkrdpdpfpvdfiapgsmiikhpmdfstmk 197

```

QY 183 EKIKNNDYOSTEELKDNPKLMCTNAMIYKNETIYYKAACKLHSGKIIISOERISLQ 242
Db 198 EKlNndygsleelkdnfklmctnamlynkpetliyykaakkllhsgmklisgerislkq 257
QY 243 SIDFMADLOKTRKQKDGTDTSQSGEDGCMQREREDSGDAEAHAFKSPSENKKKDKML 302
Db 258 sidfmalqtrkqkdgtcdsgsedgscwqredsgdeahafkspsenkkkdkdml 317
QY 303 EDKRSNNLEREBOQLDRIVKESGKILTRLVNSQCFEERRKPDGTTLLGHPVDPIYG 362
Db 318 edkrsnnlereboqldrivkesgkiltrlvnsqceferkpdgttllghpvdpiyg 377
QY 363 EPGCYRLIGMTTGRLOGSVNTLOGFEKEDKRNKVTPLYLYNIGYSSAYAHYSTFANIS 422
Db 378 epgcyrligmctgrlogsvntlogfედkdrnkvtplylynigysayahystfanis 437
QY 423 KDDSDLIYSTYGEDSDLPSPFSIHFLATCQDPYVWADLLDVLTRKGHSRTLOEWEMS 482
Db 438 kddsdliystygedsdlpspfshflatcdqpyvwadlldvltrkghsrtlgemems 497
QY 483 LPDEGHTRLDTCRKEQITTEVEPPERLDSSTODRLIAKAVTNFQVPEVDFSEAEI 542
Db 498 lpedeghtrltdakemeqitleveppgrlidsstqdrliakavtnfgvpvdfseaei 557
QY 543 FOKKIDETTRLRLRLQPAONERLSTRPPGMICLIGPS 580
Db 558 fokkideetrlrlrelgeaonerlstrppmcllignps 595

RESULT 3
ID AAU16626 standard; protein; 718 AA.
XX AAU16626:
AC
XX 07-NOV-2001 (first entry)
XX
DE Human novel secreted protein, Seq ID 1579.
XX
XX
KW Human; immunosuppressive; antiarthritic; antirheumatic;
KW cytosolic; cadherent; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virocidic; fungicide; ophthalmological;
KW vulnerary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
XX
PN MO200155322-A2.
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01341.
PE
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 18-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234998.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 13-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 PI
 XX MPI: 2001-488783/53.
 DR N-PSDB; AAS26613.
 XX
 PT New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 XX
 PS Claim 11; SEQ ID NO 1579; 980bp; English.
 XX
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders

CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention.

Query Match 96.7%; Score 2972.5; DB 22; Length 718;
 Best Local Similarity 99.1%; Pred. No. 2.2e-239;
 Matches 573; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 3 KKKKKKSDKHLVEYEVKPKLVKYGNEVTELSGSSGSDSLPEDNDHKDKRK 62
 DB 70 KKKKKKSDKHLVEYEVKPKLVKYGNEVTELSGSSGSDSLPEDNDHKDKRK 129
 QY 63 RKKRRKGKQIPGEKGRKRRVKKDRDRVENEAEKDLQCAPVRLDPERKPLT 122
 DB 130 RKKRRKGKQIPGEKGRKRRVKKDRDRVENEAEKDLQCAPVRLDPERKPLT 189
 QY 123 SLAKOEVEQTPVQALNQLMROLQRDPAFSPFYTPFIAGYSMTIKHPWDFSTMK 182
 DB 190 SLAKOEVEQTPVQALNQLMROLQRDPAFSPFYTPFIAGYSMTIKHPWDFSTMK 249
 QY 183 EKIKNDYQSIIEELKDNPKLMCTNAMIVNKPEITLYYAAKKLHSGMKLISOERISLKO 242
 DB 250 EKIKNDYQSIIEELKDNPKLMCTNAMIVNKPEITLYYAAKKLHSGMKLISOERISLKO 309
 QY 243 SIDPMADIQTRKOKDGTDSQSSEGGCWORENEDSDAFAHFKSPSKNKKDDML 302
 DB 310 SIDPMADIQTRKOKDGTDSQSSEGGCWORENEDSDAFAHFKSPSKNKKDDML 369
 QY 303 EDKFKSNLREBEOBLDITVESGKLTFRRLVNSOCFERKPRDGTTLGLHVPDPIVG 362
 DB 370 EDKFKSNLREBEOBLDITVESGKLTFRRLVNSOCFERKPRDGTTLGLHVPDPIVG 429
 QY 363 EPGYCLVRLGNTGRLOSQVNTLQGFKEKDKNNKVTPLYLYNGPYSYAPHYDSTFANIS 422
 DB 430 EPGYCLVRLGNTGRLOSQVNTLQGFKEKDKNNKVTPLYLYNGPYSYAPHYDSTFANIS 489
 QY 423 KDDSDLIYSTYGEDSDLPDSFSIHFLATCODPYVWADSLDVLTKGSHRTOEMEMS 482
 DB 490 KDDSDLIYSTYGEDSDLPDSFSIHFLATCODPYVWADSLDVLTKGSHRTOEMEMS 549
 QY 483 LPDEGHTRTLDTGKEMQITEVEPPGRIDSSTODRLTALKAVTNFGVVEVPDSEAEI 542
 DB 550 LPDEGHTRTLDTGKEMQITEVEPPGRIDSSTODRLTALKAVTNFGVVEVPDSEAEI 608
 QY 543 FQKKIDETTRLLRELQEAQNERLSTRPGNMICILGPS 580
 DB 609 FQKKIDETTRLLRELQEAQNERLSTRPGNMICILGPS 646
 RESULT 4
 AAM93712
 ID AAM93712 standard; Protein; 405 AA.
 AC AAM93712;
 XX 06-NOV-2001 (first entry)
 DT Human polypeptide, SEQ ID NO: 3652.
 DE Human polypeptide, SEQ ID NO: 3652.
 XX Human; full length cDNA; cDNA synthesis; oligo-capping.
 XX Homo sapiens.
 OS Homo sapiens.
 XX
 PM EPI130094-A2.

Db 121 iskdssdlystygedsdldpsdisihflatcdqdyymadsildivlkgshstlqeme 180
 QY 481 MSIPEDSGHRTLTPTGKMEQITEVEPPGRLDSTQDRLLAKAVTNFGVPEVEFDESEA 540
 Db 181 mslypdeghrtltclakeme-iteveppgrldstqdrllalkvtnfygvpevfdsaea 239
 QY 541 EIFQKIDETTRLRLRELOAQNRLSTRPGNMICLLGPS 580
 Db 240 eifqkldettrllrrelqaeqnerlstrppnmicllgps 279
 RESULT 6
 AAM38835
 ID AAM38835 standard; Protein: 597 AA.
 AC AAM38835;
 XX
 DT 22-OCT-2001 (first entry)
 DE Human polypeptide SEQ ID NO 1980.
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 OS Homo sapiens.
 PN MO200153312-A1.
 XX 26-JUL-2001.
 PF 26-DEC-2000; 2000WO-US34263.
 XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI: 2001-442253/47.
 DR N-PSDB: AAI57991.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 3; SEQ ID NO 1980; 10078pp; English.
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and

CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 XX
 SQ Sequence 597 AA;
 Query Match 30.0%; Score 921.5; DB 22; Length 597;
 Best Local Similarity 34.7%; Pred. No. 3.8e-68;
 Matches 213; Conservative 127; Mismatches 184; Indels 89; Gaps 17;
 QY 1 MGKKKKRHSKSD-KHLYEY-----VEKPLKVLKVGNEVELSTGSSGHDLSFEKKNH 55
 Db 1 mgkkrkhhkaewrsyedyadkplkxplkvlkvgsvevlels--gsghssyyddrdsd 58
 QY 56 --DKHKRRKKRRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKR 110
 Db 59 erehkekk 115
 QY 111 VRDLPEPKPLTSLAKQEEVEQTPLOEALNQLMRQLORRPSAFSPVTFPIAGYSM 170
 Db 116 vevppdrpractqpaenestpqllehfirgkrkdpngfifaivtdalapygsm 175
 QY 171 IIRPNDFSTYKKEIKKNDQSTIELKDNFKIMOTNMIYKPEITYYKAKKLLSGMK 230
 Db 176 ikhpmdfgtmkdklvaneysvtefkadikmednamynrptvykkakllhagfk 235
 QY 231 ILISERI-----OSLKQSIDFMADLQTRKQKQDPTDSQSGEDGCGWQREDSGDAA 284
 Db 236 mmskqgaallgnedaveepvepvq----- 262
 QY 285 HAFKSPSKENKKKDKMDLEDFK-----SNNLERQEQDLRIKESGGKLTFRVY 334
 Db 263 --vetakskkpsarevyscmfdepgnacsldstakehvalavehaeardrinfllp 319
 QY 335 NSOCEFERKRPDGTTLGLHPVPIVGEPPGYCLVRGMTTGRLOSQSVNTLOGFEKDRN 394
 Db 320 gqkngylkrngdgllysvntaepdaeeethpvdslsskllppfttl-gfkderm 378
 QY 395 KVTPLYVINYGPYSSYAPHYDSTFANISKSDSLITYSGEDSULPDSFSHFELACOD 454
 Db 379 kvt--fls-sattalsmqnsvfgdlksdemellysagdeivqcalisqetvkvkags 434
 QY 455 YPYVWADSLDLVLTFGKSGSTFLOEME-----MSLPDEGHTTRL-DYKMEQITEVEP 507
 Db 435 yskkvvddlllqitgdstlrlfqlkqrnympmpdeakvqdtlgssssvlefmsmk- 493
 QY 508 PGRLDSTQDRLLAKAVTNFGVPEVEFDESEAEITFOKKIDETTRLRLRELOAQNRLST 567
 Db 494 -----sydvsvdismsslgkvkkelddshl---nldetkllqldlneagaerzgs 544
 QY 568 RPPGMNICLLGPS 580
 Db 545 rpsnlslsnaas 557
 RESULT 7
 AAM40621
 ID AAM40621 standard; Protein: 605 AA.
 AC AAM40621;
 XX
 DT 22-OCT-2001 (first entry)
 DE Human polypeptide SEQ ID NO 5552.
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246539.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249287.
 PR 17-NOV-2000; 2000US-0249289.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-488783/53.
 DR N-PSDB; AAS26200.

XX New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 XX
 XX
 PS Claim 11; SEQ ID NO 1166; 980pp; English.

XX The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunoassay assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi

CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention.

Query Match 26.0%; Score 798.5; DB 22; Length 199;
 Best local Similarity 41.1%; Pred. No. 1.4e-58;
 Matches 176; Conservative 0; Mismatches 1; Indels 251; Gaps 1;

QY 17 EYVEKPLKLVKVGMEVTELTSGSSGHSLSFEDKNDHKDKRRKKRKGKQIPGE 76
 Db 1 EYVEKPLKLVKVGMEVTELTSGSSGHSLSFEDKNDHKDKRRKKRKGKQIPGE 60
 QY 77 EKGRKRRVVEDKKRDROVNEAEKDLQCHAPVRLDLPPEKPLTSLAKQEEVEQTP 136
 Db 61 EKGRKRRVVEDKKRDROVNEAEKDLQCHAPVRLDLPPEKPLTSLAKQEEVEQTP 70
 QY 137 QEALNQLMRQLQKRDPSAFSPVPTDIARGYSMITKHPMDFSTMKKIKNNYQSTIEL 196
 Db 71 ----- 70
 QY 197 KDNFKLMCTNAMIYNNPETIYRKANKLLSHGKILISOERIOSKSIDFMADLOKTRKO 256
 Db 71 ----- 70
 QY 257 KDGTDTSGSEDDGCMOREDESDAEAHAFSPSKENKKDKMDLEDKKSNNLERDE 316
 Db 71 ----- 70
 QY 317 QLDRIVESGKLTRLVNSQCEFERRRKPDGTTTGLIHPVDPIVGPYCLVRLGNTTG 376
 Db 71 -----ceferrrkpdgttttllghpvpdivgpycprlsmgtg 109
 QY 377 RLQSGVNTLOGFEKEDKRNKVTPLYLYNGPYSSYAPHYDSTFANISKDSLLYSTYGED 436
 Db 110 RLQSGVNTLOGFEKEDKRNKVTPLYLYNGPYSSYAPHYDSTFANISKDSLLYSTYGED 169
 QY 437 SGLPSDF 444
 Db 170 sglpsdfs 177

RESULT 9
 ABB63028
 ID ABB63028 standard; Protein; 861 AA.

AC ABB63028;
 DT 26-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID NO 15876.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 PR 23-MAR-2000; 2000US-191637P.
 XX 11-JUL-2000; 2000US-0614150.
 PA (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL07131.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS
 PS Disclosure; SEQ ID NO 15876; 21np + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS7737-ABBS2072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

Seq Sequence 861 AA;

Query Match 22.8%; Score 699.5; DB 22; Length 861;
 Best local Similarity 27.6%; Pred. No. 2.3e-49;
 Matches 211; Conservative 127; Mismatches 219; Indels 207; Gaps 22;

QY 3 KKKKKKSD-KHLYEY-----VEKPLKLVKVGMEVTELTSGSSGHD----- 45
 Db 5 KKKKKKSD-KHLYEY-----VEKPLKLVKVGMEVTELTSGSSGHD----- 45
 QY 46 ---SSLFEDKNDHKDKRRKKRKGKQIRGEGRKRRVVEKDKRDROVNEAE 102
 Db 65 ammspvrpeelqdgqhrhkrkkskkkk---korekhhkhe-krnsrdhrdags 119
 QY 103 -----KDLOCH----- 108
 Db 120 dedmaggadaacscfapsvappaadpsqdgfsfmdddsgqrpipenlffagittcn 179
 QY 109 -----APRLD-----LP----- 116
 Db 180 spsnpcvtkrlpkrlddlimgsspnssslgsslgllgspckrplldllpsptpaga 239
 QY 117 -----PEKPLTSLAKOE-----EVEQTPLOEALNQLMRQLQKRDPSAF 155
 Db 240 nslnaltpkalaepkcpsssssgrepscvlklqkqspklnllhllrflckrpbqf 299
 QY 156 FSEPVYDFIAPGYSMTIKHPMDFSTWKEKIKNNYQSTIELDNFKLMCTNAMIYNNKPT 215
 Db 300 fawpyrddmappyslstrpmdfstmrqklddheytaletfcdfflmcenaikynhndt 359
 QY 216 IYKKAKKLLHSGMKILISOERI-QSLKOSIDFMADLOKTRKOKDGTDTSGSEDDGCMOR 274
 Db 360 vynkaakrlllygmkhlpqenlmrslkplsgymrel---tarelfelsnd-----mar 411
 QY 275 ERREDSGD-----AEAHAKSPSKENKKDKMDLEDKKSNNLERDE-----NNLERQ--EQDR 320
 Db 412 emndsdegastgaepripagleeertrtlrtenaprlthfepvydltgelllaqyn 471
 QY 321 IYKESGKLTRLVNSQCEFERRRKPDGTTTGLIHPVDPIVGPYCLVRLGNTTGRLQS 380
 Db 472 aagqakgrvnaaknhkngflrkmkgdgttllnlvkee---negpeyrvttldlgkqk 528
 QY 381 GYNTILOGFEKEDKRNKVTPLYLYNGPYSSYAPHYDSTFANISKDSLLYSTYGEDSLP 440
 Db 529 gsaqltgrvdkrnnavrtvksllyngaafasfablfsrlfslsaeevtqvlrtyglassae 588
 QY 441 SDFSIHEFLATCODYPYVWADSLDLVLTGSGHSRTL-----QEMENSLPDE 487
 Db 441 SDFSIHEFLATCODYPYVWADSLDLVLTGSGHSRTL-----QEMENSLPDE 487

Db 589 yaesllgtkd-snygtiangllidlngehsksldelynmqlhysgelekcfegee 647
 QY 488 GHTRTLDGKEMEOITTEVEPPGRLDSSITODRLIALKAVTNCPVEVPDSEAEI----- 542
 Db 648 -----eessggettaqdegeyekyknchvdfkqsglgeidvstfdmeaemksyel 701
 QY 543 ---FQKRLDETRRLRELQEAQNERLSTRPGNMICLLGPSSEK 583
 Db 702 nrmhehlsqnltllekrlvaghdlrs-qplpnhlglvqpagge 744

RESULT 10
 AAB95421
 ID AAB95421 standard; Protein; 501 AA.
 AC AAB95421;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:17823.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 KW
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PE 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 XX
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX
 XX
 PT primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 17823; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX
 SQ Sequence 501 AA;
 Query Match 22.3%; Score 684; DB 22; Length 501;
 Best Local Similarity 34.7%; Pred. No. 2e-48;
 Matches 164; Conservative 93; Mismatches 164; Indels 52; Gaps 11;

QY 130 EVEQTPLEALNQLMROLQKRDPSAFSPYTDIAAGYSMTIKHPMDFTMKRKIRND 189
 Db 19 enestpiqllehlfrlqqrkdpgffafpytdalapgysmlkhpndftmkkkyane 78
 QY 190 YQSI BELKDNFKMCTNAMYNNKPEITYYKAARKLLSGMKILSOERIQSISIDPMAD 249
 Db 79 yksvtefkadfklnchdcnamtyrpdvtyvklakillagfkmskerllaktsmsfmgd 138
 QY 250 LQKTRKQKDGTFDTSGSGEDGCMQREEREDSGDAEHAHAFKSP-----SKENKKKKDKMLIED 304
 Db 139 m-----dfsqgaallg-----nedlaveepvpevpvgyetakskskperevisc 183
 QY 305 KFK-----SNLTEREQEQLDRIVKESGKILTRLVNSQCFERRKPDGTTTGLL 354
 Db 184 mfepegnacsltdstaehvalavehaadeardrinfllpgkmgylkrngdgsllysv 243
 QY 355 HPVDPYGEPEGYCLVRGAMTTGRLQSGVNTLQGRKEDKRNKVTPLYLNTGPPSYAPHY 414
 Db 244 ntaepdadeetthpvdlsissksllpgfttl-gfkdermrvkt--fls-sattalmsqn 298
 QY 415 DSTFANISKDSDLIYSTYSGSDLPSPDSTHEFLANCOQVPYMAADSLDVLKKGHSR 474
 Db 299 nsvfgdksdemelllysaegdeivgcalslgeivkdaagsykvvdlllqqlvgdgsnr 358
 QY 475 TLOEME-----MSLPDEDEGHTRTL-DTGKEMEOITTEVEPPGRLDSSITODRLIALKAVTN 527
 Db 359 tlfdlqkrnympkdpdaekvgdclgdsessvlefmsmk-----aypdvsvdlmss 411
 QY 528 FGVEVEVDSEAEI FQKRLDETRRLRELQEAQNERLSTRPGNMICLLGPS 580
 Db 412 lqkvkkelldpddshl---nidetklllqdlheagaergysrpsnlslnas 461

RESULT 11
 AAB41780
 ID AAB41780 standard; Protein; 280 AA.
 AC AAB41780;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF1544 polypeptide sequence SEQ ID NO:3088.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vunerary; antiparkinsonian; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteoprotic; antiarthritis; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasculoprotic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.

```

XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
XX Shinketsu RA, Leach M;
XX
DR WPI: 2000-602362/57.
DR N-PSDB: AAC75989.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11; Page 2306-2307; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotrophic; vulnery;
CC antipapillary; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antirheumatic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic;
CC antidiabetic; hypotensive; dermatologic; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
CC antihypertoid; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease, to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 280 AA;
SQ

```

```

Query Match 15.5%; Score 476; DB 21; Length 280;
Best Local Similarity 32.8%; Pred. No. 2e-31.
Matches 102; Conservative 63; Mismatches 90; Indels 56; Gaps 5;
QY 70 EKOIPEKGRKRRRVEDKKRRDRVENEAE-KDQCHAPVRLDLPPEKPLTSLAKQ 128
DB 2 ekhidae-----rrkkeekkrkrerehedegeaddfdpgkvevepprpractrtg 58
QY 129 EYVQGTPLQELNQLMROLDKQKPSAFSPVYDFIAPGYSMTIKHPDSTMEKIKNN 188
DB 59 pemertlqqlhflhflqrlqkdpghfaprvdaalapysmlkhpmdgtmkdkivan 118
QY 189 DYOSIEELKDKFKMCTNAMYNNKPTIYYKAAKKLHSMKLTISOERI-----QSLKQ 242
DB 119 eykavtefkedfkimchamantynpdtvyyklakllhagfkmmyskaaalignedavee 178
QY 243 SIDFMADLQKTRKOKDCTDTSQSGEDGCMQOREREDSGAEAHAFKSPSKENKKDKMDL 302
DB 179 pvepvpvqv-----vetaakkakpstrevl 202
QY 303 EDRK-----SNNLEREQQLDRYKESGKLTRLVNSQCFEERKRDGTTTGL 352
DB 203 scmlpegnacsltdsaehvialvehaadeardlnrlfpgkngylkrngdgslls 262
QY 353 LILHPVDPIVGE 363
DB 263 vvnlaepnade 273

```

```

RESULT 12
AAB95881
ID AAB95881 standard; Protein: 233 AA.
XX
AC AAB95881;
XX
XX 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:18979.
XX
XX Human; primer: detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
XX
XX 11-JAN-2000; 2000JP-0118776.
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 18979; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification, where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAB93166 to AAB13628 and
XX AAB93633 to AAB18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAB13629 to AAB13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 233 AA;
SQ

```

```

Query Match 12.7%; Score 391; DB 22; Length 233;
Best Local Similarity 65.1%; Pred. No. 1.9e-24;
Matches 69; Conservative 19; Mismatches 18; Indels 0; Gaps 0;
QY 130 EYVQGTPLQELNQLMROLDKQKPSAFSPVYDFIAPGYSMTIKHPDSTMEKIKNN 189
DB 19 enestpqlqlhflhflqrlqkdpghfaprvdaalapysmlkhpmdgtmkdkivane 78

```



```

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR N-PSDB: AA160173.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 5948; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AA038642-AA042213) with neurotrophic,
CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localized neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilization of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1109 AA;

```

```

XX
AC AAM41018;
XX
DT 22-Oct-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 5949.
XX
KW Human; neurotrophic; immunosuppressant; cytoskeletal; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000MO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR N-PSDB: AA160174.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 5949; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AA038642-AA042213) with neurotrophic,
CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localized neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilization of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1109 AA;

```

```

Query Match      8.4%; Score 258.5; DB 22; Length 1109;
Best Local Similarity 23.5%; Pred. No. 2,3e-12;
Matches 109; Conservative 62; Mismatches 136; Indels 157; Gaps 15;

```

```

OY 11 DKHLYEYVERPLKLVKVGNEVETLSTGSSGSDSLFEDKNDKDKRKRKRKRGE 70
DB 437 dvhhppgcttrplnlygdv-----emkngvcrkessvktvrs-----tskvtkkakkak 485

```

```

OY 71 K-----QIPGEEGRKRRRVKEDKK-----91
Db 486 kalaepcavlpvcapyipqrlnrqiaqvaigrkqfverahsywlkrlsrngapllr 545
OY 92 -----RDRDRVENEAEXDLOCHAPVRLDLPPEKPLTSSLAKOE-----E 130
Db 546 rlgsslgqsrssqgrendeemkaakekikwyqrlrhdlerarlllellrkreklkregvk 605
OY 131 VEO-----TPEALNQLMRQLOKDPDSAFSESPVTDFTAPGYSMTIKHPMDFSTIKE 183
Db 606 vegvamelrlplvlrsvldqldkqparlfagpvsikevdpdyldhikhpmdfatmrk 665
OY 184 KIKNDYOSIEELKDNFKIMCTNAMITYKKAATLHSGMKILSQERIOSLKOS 243
Db 666 rleaggyknlhefeedfaldldncmkynardtvyraavrlrdggvvlrqar-----718
OY 244 IDFMADIQTRKQKDGTDTSOSGEDGCGWOREREDSGDAAE--HAFKSPSEKNNK----296
Db 719 -----revdsigleasgmhlperpaaprrpfsw 748
OY 297 KDKDMLEDKFKSNL---EREOQDRI---VKESGKILR--RLVNSQCEFERKPDG 347
Db 749 edvdrllpnanrahlgleeqrlreilldmlclcamkssgsrsktrakllkelaallink--805
OY 348 TTTTGLHPVDPVIGBPQYCLVRLGTTGRLQSGVNTLQGFKEKED 391
Db 806 ---lsqghs-qplptgpg-----legfeed 826

```

Search completed: July 11, 2002, 15:49:34
 Job time: 1023 sec

OM of: US-09-687-230-1 to: SwissProt_40.* out_format : pfs

Date: Jul 11, 2002 3:48 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame+22p_model -DEV=xlp
-O=/sgn2.1/USPRO_spool/US09687230/runat_11072002_144409_29428/app_query.fasta_1.2401
-BB=SwissProt_40 -OFMT=fastan -SUFFIX=isp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORMEXT -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09687230 @CGNL.1.105 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-687-230-1
Query length: 2307
Database: SwissProt_40.*
Database sequences: 105224
Database length: 38119550
Search time (sec): 60.480000
```

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
SwissProt_40:BRP3_HUMAN	266.00	266.90	3.0e-08	1214	Q9UL44 homo sapiens (human)
SwissProt_40:BRD1_HUMAN	265.50	264.44	4.6e-08	1058	O95696 homo sapiens (human)
SwissProt_40:BRD1_HUMAN	245.50	244.73	4.5e-07	1214	P55201 homo sapiens (human)
SwissProt_40:CBP_MOUSE	244.00	238.32	5.7e-07	2442	O92799 homo sapiens (human)
SwissProt_40:CBP_MOUSE	227.50	221.28	5.1e-06	2441	P45481 mus musculus (mouse)
SwissProt_40:T2D1_DROME	217.50	212.35	1.9e-05	2068	P31123 drosophila melanogaster
SwissProt_40:P300_HUMAN	216.00	209.50	2.3e-05	2414	O99472 homo sapiens (human)
SwissProt_40:SN22_HUMAN	210.00	206.83	5.0e-05	1586	P51531 homo sapiens (human)
SwissProt_40:YC40_HUMAN	207.50	210.73	6.6e-05	733	O91011 homo sapiens (human)
SwissProt_40:SPT7_YEAST	201.50	199.52	0.0002	1332	P35177 saccharomyces cerevisiae
SwissProt_40:ATRX_CAEEL	198.50	196.25	0.0002	1359	O90759 caenorhabditis elegans
SwissProt_40:SN24_HUMAN	194.00	189.95	0.0004	1647	P31552 homo sapiens (human)
SwissProt_40:T2D1_HUMAN	184.50	176.01	0.0022	1872	P21675 homo sapiens (human)
SwissProt_40:BRD2_HUMAN	180.50	182.10	0.0024	801	P25440 homo sapiens (human)
SwissProt_40:TRDN_HUMAN	170.00	172.06	0.0094	728	O10611 homo sapiens (human)
SwissProt_40:BDP1_YEAST	168.50	171.01	0.0115	686	P35817 saccharomyces cerevisiae
SwissProt_40:CHM_DROME	167.50	161.07	0.0142	1982	O97159 drosophila melanogaster
SwissProt_40:BRD4_HUMAN	167.00	163.70	0.0147	1362	O60885 homo sapiens (human)
SwissProt_40:NEST_HUMAN	166.00	161.23	0.0171	1618	P48681 homo sapiens (human)
SwissProt_40:USO1_YEAST	164.50	158.83	0.0210	1790	P25386 saccharomyces cerevisiae
SwissProt_40:MYH9_DROME	163.00	158.02	0.0254	1638	P25439 drosophila melanogaster
SwissProt_40:MYH9_DROME	163.00	153.42	0.0383	1959	P14105 gallus gallus (chick)
SwissProt_40:GAPR_PLAUF	159.50	161.82	0.0377	678	P13816 plasmodium falciparum
SwissProt_40:Y1L7_CAEEL	159.50	157.53	0.0393	1130	O11102 caenorhabditis elegans
SwissProt_40:PCNT_MOUSE	158.50	153.08	0.0409	1920	P48725 mus musculus (mouse)
SwissProt_40:MAB3_MOUSE	158.00	149.43	0.0508	2464	P14873 mus musculus (mouse)
SwissProt_40:BRD3_HUMAN	157.50	159.18	0.0495	726	O15059 homo sapiens (human)
SwissProt_40:TRHY_RABIT	157.00	153.10	0.0556	1407	P37079 oryctolagus cuniculus
SwissProt_40:CENH_HUMAN	157.00	147.75	0.0584	2663	O02224 homo sapiens (human)
SwissProt_40:MARA_HUMAN	156.50	146.80	0.0626	2805	P78559 homo sapiens (human)
SwissProt_40:ICR1_DROME	156.00	154.16	0.0623	1097	O96433 drosophila melanogaster
SwissProt_40:WDR3_HUMAN	156.00	147.96	0.0659	2295	O95616 homo sapiens (human)
SwissProt_40:Y1L8_CAEEL	155.43	155.43	0.0655	887	P45504 caenorhabditis elegans
SwissProt_40:TOP1_CRIGR	155.00	154.07	0.0901	767	O07050 cricetus cricetus
SwissProt_40:REST_CHICK	148.82	148.82	0.0946	1433	O13184 gallus gallus (chick)
SwissProt_40:MYST_DROME	145.43	145.43	0.1037	2017	O99323 drosophila melanogaster
SwissProt_40:TOP1_MOUSE	152.00	153.04	0.1029	767	O04750 mus musculus (mouse)
SwissProt_40:TF1A_HUMAN	152.00	150.40	0.1054	1050	O15164 homo sapiens (human)
SwissProt_40:MYH8_HUMAN	152.00	145.11	0.1106	1972	P35749 homo sapiens (human)
SwissProt_40:MYH8_BOVIN	151.50	144.57	0.1182	1976	O27991 bos taurus (bovine)

seq_name: SwissProt_40:BRP3_HUMAN

ID	BRP3_HUMAN	STANDARD	PRT	1214 AA
AC	Q9UL44:			
DT	16-OCT-2001 (rel. 40, Created)			
DT	16-OCT-2001 (rel. 40, Last sequence update)			
DE	Bromodomain and PHD finger-containing protein 3 (Fragment).			
GN	BRP3 OR KIAA1286.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=20039619; PubMed=10574462;			
RA	Nagase T., Ishikawa K.-I., Kikuno R., Hirosewa M., Nomura N.,			
RA	Ohara O.;			
RT	"Prediction of the coding sequences of unidentified human genes. XV.			
RT	The complete sequences of 100 new cDNA clones from brain which code			
RT	for large proteins in vitro."			
RL	DNA Res. 6:337-345(1999).			
CC	-1- SIMILARITY: CONTAINS 1 BROMODOMAIN.			
CC	-1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.			
CC	-1- SIMILARITY: CONTAINS 1 PMP DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: AB03112; BAA86600.1; -			
DR	InterPro: IPR001487; Bromodomain.			
DR	InterPro: IPR001965; PHD.			
DR	InterPro: IPR000313; PMP.			
DR	Pfam: PF00439; bromodomain; 1.			
DR	Pfam: PF00628; PHD; 1.			
DR	Pfam: PF00855; PMP; 1.			
DR	PRINTS: PR00503; BROMODOMAIN.			
DR	SMART: SM00297; BROMO; 1.			
DR	SMART: SM00249; PHD; 2.			
DR	SMART: SM00293; PMP; 1.			
DR	PROSITE: PS00633; BROMODOMAIN_1; FALSE_NEG.			
DR	PROSITE: PS00614; BROMODOMAIN_2; 1.			
DR	PROSITE: PS00812; PMP; 1.			
DR	Zinc-finger; Bromodomain.			
KW	Zinc-finger; Bromodomain.			
FT	NON_TER 1			
FT	ZN_FING 223 271			
FT	DOMAIN 415 441			
FT	DOMAIN 615 685			
FT	DOMAIN 1085 1168			
FT	DOMAIN 1214 AA; CA490810622109CD CRC64;			
SO	SEQUENCE			

alignment_scores:

Quality	266.00	Length	663
Ratio	0.866	Gaps	26
Percent Similarity	46.305	Percent Identity	22.775

alignment_block:

US-09-687-230-1 x BRP3_HUMAN

Align seg 1/1 to: BRf3_HUMAN from: 1 to: 1214

```

74 CGCGCCCTCGCGCTGGCCGCGGACCGGAAGCGGCGCGCGCGCTGCG 123
   |||||
397 ProPoglyAlaIaIaThrAlaIaIaArgIaIaSerProAlaIaIaSerI 413
   |||||
124 GCGTCGCGCGGCGGCGGACCGGCGCGCTCGACATGGG..... 166
   |||||
413 eSerIuIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 430
   |||||
167 .....CAAGAAGCAAGAAGCAAGAAGCAAGAAGCAAGAAGCAAGAAG 196
   |||||
430 IuIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 446
   |||||
197 ACA...CGCTACAGAGAGTA..... 214
   |||||
447 SerIuIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 463
   |||||
215 .....TGTAAGAAGCGCTGAAGCTGGT..... 238
   |||||
463 mIuIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 480
   |||||
239 .....CGTCAAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 271
   |||||
480 euProMetIeuaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 496
   |||||
272 CTCCAGCGGCGAGCTCGGCGGACGAGCTCCAGCTCTTCGAAGCAAGAGAG 321
   |||||
496 eCySerIeIuSer.....PheGlnaIaIaIaIaIaIaIaIaIaIaIaIaIa 507
   |||||
322 ATCATGACAAACACAAAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAG 371
   |||||
507 IuPheMetIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 523
   |||||
372 AAGCAGATTCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 421
   |||||
524 AaNgIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 540
   |||||
422 GGATAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 471
   |||||
540 gAsnIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 557
   |||||
472 ATCTCAGAGTCAAGCGGCTGTGAGATTGACTGCTCTCGAGAGAGCT 521
   |||||
557 IuIeuaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 573
   |||||
522 CTACAGAGCTCTTTAGCCAAACAGAGAGAGTA.....GAACAGAGAGAG 565
   |||||
574 IeuiIeIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 590
   |||||
566 CCTTCAAGAGAGCT.....TTGAATCAAGCTGA 591
   |||||
590 sValGlnIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 607
   |||||
592 TGAGACAA.....TTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 629
   |||||
607 euArgIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 623
   |||||
630 TCATTTCCGTGAGCTGATTTATGCTCTGCTACTCCATGATCAATGA 679
   |||||
624 AaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 640
   |||||
680 ACACCAATGATTTTATGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 729
   |||||
640 rIuSPrometIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 657
   |||||
730 ATCAGTCCATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 779
   |||||
657 yArgIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 673
   |||||
780 GGCATGATTTTACATTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 829
   |||||
674 CyMeIuIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 690

```

```

830 GCTGTTCACATCAGAGATGAAATTTCTTACCCAGAGAGAGAGAGAGAGAG 879
   |||||
690 gIeuaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 707
   |||||
880 TGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 929
   |||||
707 IuAsnIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 713
   |||||
930 AAAGATGAGACAGACACCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 979
   |||||
714 GlnaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 730
   |||||
980 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
   |||||
730 gPheSerIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 747
   |||||
1021 TCAGAGATCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1070
   |||||
747 IuSerIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 753
   |||||
1071 GATAGTTTAAAGCAATATTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1120
   |||||
754 .....LeuIuIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 763
   |||||
1121 CATC.....GTGAAGAGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1161
   |||||
763 uValSerIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 780
   |||||
1162 TGACAGTCAAGTCCGATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1211
   |||||
780 euArgIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 790
   |||||
1212 TTGGAGCTCTCCATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1261
   |||||
791 .....LeuAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 797
   |||||
1262 CCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1311
   |||||
797 ..... 797
1312 CTTTCAGAGGAGTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1361
   |||||
798 .....GlnProIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 809
   |||||
1362 TATTTGAATTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1411
   |||||
810 GlnIeuaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 825
   |||||
1412 ATTTGCAATATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1461
   |||||
825 aIeuaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 836
   |||||
1462 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1511
   |||||
836 rGAspIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 844
   |||||
1512 AGCTCCAGAGATTTATCGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1561
   |||||
845 .....ProIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 847
   |||||
1562 AACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1611
   |||||
847 uIuProIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 863
   |||||
1612 CTGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1661
   |||||
863 rGlnIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 869
   |||||
1662 CAGATTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1711
   |||||
870 ProIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 879

```



```

1712 CAGGCTCATAGAGCGCTGAAGACACTACAAATTTGGCGTTCAGCTGAG 1761
880 .....PheLeuLysProArgLysV 886
1762 TTTTTCAGCTGTGAAGAAGCTGAATATATTCACAGAA..... 1799
886 aL.....GlulIuAspGlulLeuLugLysSerProLeuInLeu 899
1800 ...CTTGATGACAGCACACAGATTGCTTCAGGAACCTCAGAGACCCAGAA 1846
900 GlysAngLupProLeuGlnArgLeuLeuSerAspAngLylLeaSnArgL 916
1847 TGAACGTTGAGCACCAGAC...CCCGTGGAACA 1878
916 euserLeuMetaIaProAspThrProAlaGlyThr 927

seq_name: SwissProt_40:BRD1_HUMAN

seq_documentation_block:
ID BRD1_HUMAN STANDARD; PRT; 1058 AA.
AC O95696;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bromodomain-containing protein 1 (BR140-like protein).
GN BRD1 OR BR1 OR BRPF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=20071128; PubMed=10602503;
RA McCallagh P., Chaplin T., Meerabux J., Grenzelias P., Lillington D.,
RA Poulson R., Gregorini A., Saha V., Young B.D.,
RT "The cloning, mapping and expression of a novel gene, BR1, related to
RT the APL1 leukaemia gene.";
RT Oncogene 18:7442-7452(1999).
[2]
SEQUENCE FROM N.A.
RX Hunt A.;
RA Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RL -1 SUBCELLULAR LOCATION: Nuclear.
CC -1 TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS.
CC -1 SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -1 SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC -1 SIMILARITY: CONTAINS 1 PWWP DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
DR EMBL; AF005067; AAP34320.1; -
DR EMBL; Z98885; CAB11574.1; -.
DR MIM; 604589; -
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR001965; PHD.
DR InterPro; IPR000313; PWWP.
DR Pfam; PF004439; bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR Pfam; PF00855; PWWP; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 2.
DR SMART; SM00293; PWWP; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS50812; PWWP; 1.
DR NCBI; protein; Zinc-finger; Bromodomain.

```

FT	ZN_FING	214	264	PHD-TYPE.
FT	DOMAIN	579	649	BROMODOMAIN.
FT	DOMAIN	929	1012	PWWP.
50	SEQUENCE	1058 AA;	119519 MW;	6E7B07E8A030E104 CRC64;
<p>alignment_scores:</p> <p>Quality: 262.50 Length: 485</p> <p>Ratio: 1.156 Gaps: 18</p> <p>Percent Similarity: 46.804 Percent Identity: 23.918</p>				
<p>alignment_block:</p> <p>US-09-687-230-1 x BRDL_HUMAN ..</p>				
<p>Align seg 1/1 to: BRDL_HUMAN from: 1 to: 1058</p>				
17	CGCGCCGGCGCGCGCCGCCCTGCTCG.....	46		
390	ProProGlycylThrArgArgProLeuAsnIleTyrGlnValGluMe	408		
47	CGGGCGGGGGCTCGCGGGCGCGCTCCGCCCTCGCTCGCGCTG	95		
406	tylssnGlnValCysArgGlys.....GluSerValIlysthrValA	421		
96	GGACCGGAAGCGGGCGCGCACAGCGCTGGCGTGGCGGGGGCGG	143		
421	rGSerThrSerIlyValArgIlyslYslYslYslYslYslYslYsl	437		
146	CGGGGGCCGGTCCGCGCATGGGCGAAGACACAAAGCAAGCATGG	195		
438	AlaGluProCys.....	441		
196	AACACCTCTACAGAGAGATGT.....	217		
442	AlaValIleuProthrIValCysAlaProTyrIleProGlnArgLeuA	458		
218	..AGAGAAAGCCCTGGAAGCTGGCTCCTCAAGTAGAGAG.....	256		
458	snArgIleAlaAsnGlnValAlaIleGlnArgIlyslGlnPheValGlu	474		
257	CGAATCCACCGA.....ACHTCCACGGGGAGCTGGGGACAGAC	303		
475	ArgAlaIleHisSerTyrTrpLeuLeuIlyArgLeuSerIrnGlnIlyAla	491		
304	TCCTTGAAGACAAAAAGCATCATGACAAACAAAGACAGAAAGCGGAAA	353		
491	oleuLeuIrnArgLeuGlnSer..serLeuGlnSerGlnArgSerSerGln	507		
354	AGACAAAGAAAAGAGAGAACACATTCACGGGAGAAAAGGGAAGAAA	403		
508	GlnArgGlnAsnAspGluGlnMetCysAlaIleCysIlyslYslYslY	524		
404	ACGGAGAGAGATTAAAGAGATATAAAAGAGAG.....	434		
524	rTrpIlnArgLeuIrnArgHisAsnLeuGlnArgAlaIrnLeuIleGluL	541		
435CGAGATCGAGACCGGGTGGAGATGAGCGCAAAAAGATCTCCAG	479		
541	euleuIrnArgIlyArgIlyslYslYslYslYslYslYslYslYslGluGln	557		
480	TGTCACGGCCCTGTGAGATTAGACTTGCTGCTCTGAGAAAGCCTTCACAG	529		
558	ValAlaMetGluLeuIrnArgLeu.....	564		
530	CTCTTTAGCCAAACAGAAAGATGAAACACACACCCCTTCAAGAGC	579		
565ThProLeuThrValIleuL	571		
580	TGAATCAACTGTGTGACAAATTTGACAGAAAGATCAAGTCTTCTTT	629		
571	eulArgSerValIleuAspGlnIleuGlnAspIlyslYslAspProIlaArgIlePhe	587		
630	TCATTTCCTGTGACTGATTTTATTGCTCTGTGCTACATCCATGATCATTA	679		

```

588  AAlaGlnProValSerLeuValGluValProAspTyrLeuAspHisIleLeu 604
680  AACACCAATGATTTTACATGATGAAGAAAGATCAAGAACATGACT 729
604  SHSPrometAspPheAlaThrMetArgSarGluGluAlaGlnGlyT 621
730  ATCAGTCATAGAACAACTAAAGATTAAGTCAACTAATGCTTACTAAT 779
621  yrlYsAsnLeuHisGluPheGluGluAspPheAspLeuIleLeuAsp 637
780  GCCATGATTTACATTAACCCAGACACTTATTTAAGCTGCAAGAA 829
638  CysMetIlyTyrAsnAlaArgAspThrValPheTyrArgAlaAlaVal 654
830  GCGTTCGACATGCAAGTAAATCTTAGCCAGAAAGATTCAGAGCC 879
654  GluAlaGAspGlnGlyGlyValValLeuArgGlnAlaValArg 667
880  TGAAGCAGACATAGACTTCATGGCTGACTTCAGAAATCGAAACAG 929
667  ..... 667
930  AAAGATGACAGACACCTCAGAGATGGGAGAGAGGCGCTGGCGCA 979
667  ..... 667
980  GAGAGAGAGAGAGACTTGAGATGCCGAGCA.....CAGCGCT 1020
668  ArgGluValAspSerIleGlyLeuGluGluAlaSerGlyMetHisLeu 684
1021  TCAGAGTCCCGCAAGAAATAAAG.....AAGACAAA 1058
684  rodLurProAlaAlaAlaProArgArgProPheSerTyrGluAspVal 700
1059  GATATCTTGAGATAGTTTAAACCAATTA.....GAGAG 1099
701  ASPATGLeuAspProAlaAsnArgAlaHisLeuGlyLeuGluGlu 717
1100  AGAGCAGAGCAGCTTGACCGCATC.....GTGAAAGAA 1134
717  nleuArgGluLeuLeuAspMetLeuAspLeuThrGlySalMetIlySer 734
1135  CTGGAGGAAAGCTGACCAAG...CGGCTTGACAGCTAGTCCGAATTT 1181
734  ecIySerIySerIyArgAlaIyLeuIyGlyIyLeuIyAlaIyLeu 750
1182  GAAAGAGAAACCAAGATGGAACAGACAGCTTGCACTTCATCTGCT 1231
751  LeuArgAsnIlyS.....LeuSerGlnGlnHisSer.. 760
1232  GGATCCCATTTAGAGAGACCGACTACTGCGTGTGAGACTGGGAATGA 1281
761  GlnProLeuProThrGlyProCly..... 768
1282  CAACCTGGAAGACTTCAGTCTGGAGTCAATACTTTCAGAGGCTTCA 1331
769  .....LeuGluGlyPheGluGlu 774
1332  GAT 1334
775  Asp 775
seq_name: SwissProt_40:BRFL_HUMAN
seq_documentation_block:
ID BRFL_HUMAN STANDARD: PRT: 1214 AA.
AC P55201.09UHD;
DT 01-OCT-1996 (rel. 34, Created)
DT 01-OCT-1996 (rel. 34, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Peregrin (Bromodomain and PHD finger-containing protein 1) (BR140
DE protein).

```

```

GN BRPFI OR BR140.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94161726; PubMed=7906940;
RA Thompson K.A., Wang B., Argraves W.S., Glancottl F.G., Schranck D.P.,
RA Ruoslahti E.;
RT "BR140, a novel zinc-finger protein with homology to the TAF250
RT subunit of "TIFID";
RL Biochem. Biophys. Res. Commun. 198;1143-1152(1994).
RP [2]
RP SEQUENCE FROM N.A.
RA Hu S.N., Dong W., Zeng Y.X., Yu J., Yang H.M.;
RL Submitted (Aug-1999) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: UNKNOWN. POSSIBLE TRANSCRIPTION ACTIVATOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- TISSUE SPECIFICITY: HIGH LEVELS IN TESTIS.
CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC -1- SIMILARITY: CONTAINS 1 PWWP DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M91585; AB02119.1; -.
DR EMBL; AF176815; AAF19605.1; -.
DR MIM; 602410; -.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR001965; PHD.
DR InterPro; IPR000313; PWWP.
DR InterPro; IPR000822; Znf-C2H2.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR Pfam; PF00855; PWWP; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 2.
DR SMART; SM00293; PWWP; 1.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS50812; PWWP; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein;
KW Zinc-finger; Bromodomain.
FT ZN_FING 21 47 C2H2-TYPE.
FT ZN_FING 273 323 PHD-TYPE.
FT ZN_FING 386 400 C4-TYPE.
FT ZN_FING 645 715 BROMODOMAIN.
FT DOMAIN 1085 1168 PWWP.
FT CONFLICT 299 299 E -> A (IN REF. 2).
FT CONFLICT 729 729 V -> L (IN REF. 2).
SQ SEQUENCE 1214 AA; 137542 MW; C530CD2F3083A53D CRC64;

```

alignment_scores:

```

Quality: 245.50 Length: 347
Ratio: 1.240 Gaps: 12
Percent Similarity: 57.061 Percent Identity: 27.089

```

alignment_block:

```

US-09-687-230-1 x BRFL_HUMAN ..
Align seg 1/1 to: BRFL_HUMAN from: 1 to: 1214

```

[illegible]

```

1131 GAATCTGAG6AAGAAAGTGAACGACGGCGGCTGTGGAACAGTCAGCAATT 1180
      ..... ||| |||||..... ||| |||
814 LysGIuMetThrAlaLeuArgArgLysIleuAlaHis...GlaArgGIuTh 829
      ||| |||::: ||| :::: ||| |||
1181 TGAAGAACAAGAAAAACAGAT.....GGAACAACAGACTGTGGCAGCTCTCC 1224
      ||| |||::: ||| :::: ||| |||
829 rglArGApArgglyProGluArgHIsLylgProSerSerArgGlySerLeut 846
      ||| |||::: ||| :::: ||| |||
1225 ATCTGTGCATGCCATGTAGACAGCCAGGCTACTGCG 1262
      ||| |||::: ||| :::: ||| |||
846 hrPro.....HisProAlaAlaLacys 852
seq_name: SwissProt_40:CBP_HUMAN

seq_documentation_block:
ID CBP_HUMAN STANDARD; PRT; 2442 AA.
AC Q9Z793; Q16376; O00147;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DI 01-MAR-2002 (Rel. 41, Last annotation update)
DE CREB-binding protein.
GN CREBBP OR CBP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97385172; PubMed=9238046;
RA Schoplo O.M., Borrow J., Tomek R., Reshnal S., Harden A.,
RA Schlegelberger B., Housman D., Doggett N.A., Kowley J.D.,
RA Zelezniak E.N.J.;
RT "MLL is fused to CBP, a histone acetyltransferase, in therapy-related
RT acute myeloid leukemia with a t(11;16)(q23;p13.3).";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8732-8737,(1997).
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97321049; PubMed=9177780;
RA Giles R.H., Petrij F., Dauwerse H.G., den Hollander A.I.,
RA Lushtnikova T., van Ommen G.J.B., Goodman R.H., Deaven L.L.,
RA Doggett N.A., Peters D.J.M., Breuning M.H.;
RT "Construction of a 1.2-Mb contig surrounding, and molecular analysis
RT of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome
RT 16p13.3."
RL Genomics 42:96-144,(1997).
RL [3]
RP SEQUENCE OF 1-405 FROM N.A.
RX MEDLINE=96376968; PubMed=8782817;
RA Borrow J., Stanton V.P., Andresen J.M., Becher R., Behm F.G.,
RA Hagganil R.S.K., Clavin C.I., Distchev C., Dube I., Fischaut A.M.,
RA Hoshman D., Mittelman F., Volinia S., Watmore A.E., Housman D.E.;
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
RT a putative acetyltransferase to the CREB-binding protein.";
RL Nat. Genet. 14:33-41,(1996).
CC -1- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS
CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
CC CAMP-RESPONSIVE GENES.
CC -1- SUBUNIT: INTERACTS WITH SMAD1, SMAD2 AND SMAD3.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
CC T(8;16)(P11,P13) INVOLVING CBP AND MOZ, AND T(11;16)(Q23;P13.3)
CC INVOLVING CBP AND MLL.
CC -1- DISEASE: DEFECTS IN CREBBP ARE THE CAUSE OF RUBINSTEIN-TAYBI
CC SYNDROME (RUS), A DISORDER CHARACTERIZED BY CRANIOFACIAL
CC ABNORMALITIES, BROAD THUMBES, BROAD BIG TOES, MENTAL RETARDATION
CC AND A PROPENSITY FOR DEVELOPMENT OF MALIGNANCIES.
CC -1- SIMILARITY: CONTAINS 1 BROMOOAMIN.
CC -1- SIMILARITY: CONTAINS 1 ZZ-TYPE ZINC FINGER.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its

```



```

906 .....GACTTGCAGAAACTCGAAGACAGAAAGATGAAACGACCTCA 950
      |||
1213 ystYcIlyLysGlnLeuGysThrIleProArgspAlaIaIaYrYrSer 1229
      |||
951 CAGAGTGGGAGGAGCGAGGCTC.....TGCAGAG 982
      |||
1230 TYGInAsnArgTrHisPheCysGluTysCysPheThrGluIleGlnI 1246
      |||
983 AGAGAGAGAGGACTCTGAGATCGGAGACAGACGCTTCAGAGTCCCA 1032
      |||
1246 yGluAsnValThrLeuGlyAsp.....AspProS 1256
      |||
1033 GCAGAGAAATATAAAGAAAGACAAAGATATGCTGAGATTAATTAA 1082
      |||
1256 ercInProGlnThrIleSerLysAspGlnPheGlnLysLys...Lys 1271
      |||
1083 AGCATTAATTAGAG.....AG 1099
      |||
1272 AsnAspThrLeuAspProGluProPheValAspCysLysGluCysG 1288
      |||
1100 AGAGCAGAGCAGCT..... 1115
      |||
1288 gLysMetHisGlnIleCysValLeuHisTryAspIleIleTrpProSer 1305
      |||
1116 .....GACGCATCGTGAAGGAATCTGA..... 1139
      |||
1305 LysPheValCysAspAsnCysLeuLysLysThrGlyArgProArgLys 1321
      |||
1140 .....GGAAGCTGACGAGCGGCT 1159
      |||
1322 AsnLysPheSerAlaLysArgLeuGlnThrThrArgLeuLysAsnHis 1338
      |||
1160 TGTGAACAGCAGTGCAGATTTGAAGAGAAACGATGAAACAACA 1209
      |||
1338 uGluAsnArgValAsnLysPheLeuArgArgGlnAsn..... 1350
      |||
1210 CGTTGGACTTCCTCATCTCTGGATCCATTGTAGAGACCGAGCTAC 1259
      |||
1351 .....HisPro.....GluAlaGlyGlu 1356
      |||
1260 TGCCTGGTGGAGTGGGATGACGAGTGAAGACTGAGCTGGAGTGA 1309
      |||
1357 ValPheValArgValValAlaSerSer..... 1365
      |||
1310 TACTTTCAGAGGTTCAAGAGATTAAGAAAGATGATCCTCAGTGT 1359
      |||
1366 .....AspLysThrValGluValLysProGly 1375
      |||
1360 TA.....TATTGAATTATGGCCCTACAGTCTTTAGCAGCCGAT 1400
      |||
1375 eLysSerArgPheValAspSerGlyGluMetSerGluSerPheProTyr 1391
      |||
1401 .....TATACATCCATTTGCAAAATATCAGAGAGATGATTCGTAT 1447
      |||
1392 ArgThrLysAlaLeuPheAlaPheGluIleLeuAspValIcy 1408
      |||
1448 CTAT.....TCAACTATGGGAGAGCTCTGATCTTCAT 1482
      |||
1408 sPhePheGlyMetHisValGlnGluTyrGlySerAspCysPheProPro 1425
      |||
1483 GTGAT.....TTACGATCCATGAGTTTGTG 1508
      |||
1425 snThrArgArgValTyrIleSerTyrLeuAspSerIleHisPheArg 1441
      |||
1509 GCCACGTGC.....CAAGATTATCCGATATGTCATGGCAGATAGT 1549
      |||
1442 ProArgCysLeuArgThrAlaValTyrHisGluLeuIleGlyTyrLe 1458
      |||
1550 ACTGAGATGTTTAAACAAGCA.....GGCATTTCAGAGACCTTAC 1550
      |||
1458 uGluTyrValLysLysLeuGlyTyrValThrGlyHisIle..... 1471

```

```

1591 AAGAGATGGAGATGTCATTGCTGAAGATGAGGC..... 1625
      |||
1472 .....TrpAlaCysProProSerGlnGluLysAspTyrIlePhe 1484
      |||
1626 CATATGAGACACTTGACACAGAGAAAGAAATGACAGATTACAGACT 1675
      |||
1485 HisCysHisProProAsp.....GlnLysIle 1493
      |||
1676 AGAGCCACAGCGCGCTTGGAC..... 1697
      |||
1493 eProLysProLysArgLeuGlnGluTrpTyrLysLysMetLeuAspLys 1510
      |||
1698 .....TCCACT 1703
      |||
1510 LapeAlaGluArgIleIleHisAspTyrLysAspIlePheLysGlnAla 1526
      |||
1704 ACTCAGACAGCGCTCATAGCGTGAAGACGTAACAAATTTGCGCTTC 1753
      |||
1527 ThrGluAspArgLeuThrSerAlaLysGluLeuProTyrPhe..... 1540
      |||
1754 AGTTGAAGTTTTGACTCTGAGAGAGCTGAATATTCAGAGAAACTTG 1803
      |||
1541 .....GluGlyAspPheTrpProAsnValLeuG 1550
      |||
1804 ATGAGACCAACCAATTGCTCAGGAGACTCCAGAGACCCAGATGAAGT 1853
      |||
1550 LysLysSer.....IleLysGlnGluGlnGlnGlnGlnGluArg 1563
      |||
seq_name: SwissProt_40:CBP_MOUSE
seq_documentation_block:
ID CBP_MOUSE STANDARD: PRT; 2441 AA.
AC P45481;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE CREB-binding protein.
GN CREBBP OR CBP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=94019666; PubMed=8413673;
RA Christia J.C., Kwok R.P.S., Lamb N., Hagiwara M., Montminy M.R.,
RA Goodman R.H.;
RT "Phosphorylated CREB binds specifically to the nuclear protein CBP.";
RL Nature 365:855-859(1993).
CC -!- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP ADJUNCTS
CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
CC CAMP-RESPONSIVE GENES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -!- SIMILARITY: CONTAINS 1 ZN-ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S66385; AAB28651.1; -
DR TRANSFAC: T01318; -
DR MGD: MGI:1098280; Crebbp.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR003101; KIX.
DR InterPro: IPR000197; TAZ_finger.
DR InterPro: IPR000433; ZnF_Z2.

```



```

1081 AAGCAATTAATTAGAG..... 1097
1272 ysasnasprtlleuasprrogluprophevalaspcysylsglucysgly 1288
1098 AGAGACAGACGACGCT..... 1115
1289 Arglymethisglinilecysvalleuhslyrpsleltetprowse 1305
1116 .....GACCGATCGTGAAGAACTGGA..... 1139
1305 rcllyphevalcysaspasncysleuylsyrthrglyarproarglysg 1322
1140 .....GAAAGCTGACCGCGG 1157
1322 lualslysrheserlalyargleuglntlrthargleugllyasnhis 1338
1158 CTTGTGACAGATCGATCGCAATTTGAAAGCAAAACGATGAAACAC 1207
1339 leuglupasrlyvalasnllysrheleuarglgnasn..... 1351
1208 GACGTGGACATTCCTCCATCTGTGATCCCATTTGAGAGACGACGCT 1257
1352 .....HisPro.....GluAlaGlyG 1357
1258 ACTGCTGTGAGACGAGGAGATGACACCTGCAAGACTTCAGTCTGAGTIG 1307
1357 lualrphervalrlyvalalaserSer..... 1366
1308 AATTAATTTGACAGGGTTCAAGAGATAAAGCAAAAGTCACTCCAGT 1357
1367 .....AspLysThrValGluValLysProG 1375
1358 GTTA.....TATTTGAATTAATGCGCCCTACAGTCTTATGACCGC 1398
1375 ymetLysSerargrphervalaspsrclglylmetSerGlnSerPhePro 1392
1399 AT.....TATGACGCC 1409
1392 yArgrThrlyalaleuherlealrpeglnlgluileasrcllyalalrval 1408
1410 ACATTTGCAATATACAGCAAGGATTTCTGATTTATC..... 1448
1409 CysrPhearglymethisvalglinsrphrleuilealrprohlscl 1425
1449 .....TATTCACCTATGCGGAGACCTGATCTTCCAA 1482
1425 nlllelnllycysvaltyrllleSetyLleuasp..... 1436
1483 GTGATTTGACATCCATGAGTTTGGCCACGCTG..... 1517
1437 .....SerlleHisrPheherlealrProalrGcysleuargThrAlaVal 1450
1518 .....CAAGATTATCCGTA 1531
1451 tyrhlsgluileuilegllytyrleuglntlyrvalylslysleuValty 1467
1532 TGTCATGCGCATAGTTACTGAGATGTTTAAACAAGGAGG..... 1574
1467 rValThrAlahlsletlrpalacysproproSerclnlellyasrpyr 1484
1575 .....CATTCAGACACCTTA 1589
1484 lePheHlsCysHlsrProproasrlnllyslleProlysrProlysrArgLeu 1500
1590 CAAGAGATGAGATGTCATTCCTCGAAGATGAGCCATCTAGACACT 1639
1501 GlnGluTrrpyrLysLysMetleuAspLysAlarPhealaglualrGlel 1517
1640 TGACAGAGAAAGAAATGAGACGATTAACAGAGTAGAGCACCAGGCG 1689
1517 easnasrpyrlyasrple..... 1523

```

```

1690 GTTTGACCTCCAGTACACAGACGGCTCATAGCGCTTAAGACAGTACA 1739
1524 ..PheLysGlnAlasmsnclnasprlyleuThrSerAlaLysGlnLeuPro 1539
1740 AATTTGGCGTTCAGATTGAAGTTTTCAGTCTGAGAGAGCTGAATATAT 1789
1540 TyrPhe.....GluGlyAsrPheTr 1546
1790 CCAGAAAGAACTTGATGAGACACCATGCTCAGAGAACTCCAGAAAG 1839
1546 ProasnValleuGlnGlnSer.....IleLysGlnleuGlnG 1560
1840 CCAGAAATGACGT.....TGAGCACACGACCC 1868
1560 lueGlnGlnlueGlnLysGlnleuSerThrAlaAlaSerGluThrPro 1576
1869 CTTGGGAACATGATCTGTCTTGGTCCCTCATCAGAGAAATGCATCTT 1918
1577 GluGlySerGln.....GlyAspSer..... 1583
1919 GCTGACAAAGTACCAATTAATCTTAAGAAATGACACGCAAGTACTCCA 1968
1584 .....LysAsnAlaLysLysLysAsnAsnL 1592
1969 GGTGATATCGTAACAGATATGAGT...TCGAAAGCAATGGGAGATTTC 2015
1592 yslYsthrAsnLysAsnLysSerSerlleSerAlrGlnAlaAsnLysLys 1608
2016 CATTCCTTCCCGCTCATGGAACAAACTT 2045
1609 ProserMetProAsnValSerAsnAspLeu 1618

seq_name: SwissProt_40:TP2D1_DROME
seq_documentation_block:
ID TP2D1_DROME STANDARD; PRT; 2068 AA.
AC PS1123;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription Initiation factor TFIID 230 kDa subunit (TAFII-230)
GN (TAFI1250) (TBP-associated factor 230 kDa) (P230).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 63-75 AND 540-546.
RX MEDLINE=93279463; PubMed=8504928;
RA Kokubo T., Gong D.-W., Yamashita S., Horikoshi M., Roeder R.G.,
RA Nakatani Y.;
RT "Drosophila 230-kD TFIID subunit, a functional homolog of the human
RT cell cycle gene product, negatively regulates DNA binding of the TATA
RT box-binding subunit of TFIID."
RL Genes Dev. 7:1033-1046(1993).
CC -!- FUNCTION: MAY PLAY AN ESSENTIAL ROLE IN TFIID ASSEMBLY BY
CC INTERACTING WITH BOTH TBP AND OTHER TAF. AS WELL AS SERVING TO
CC LINK THE CONTROL OF TRANSCRIPTION TO THE CELL CYCLE. ESSENTIAL FOR
CC PROGRESSION OF THE G1 PHASE OF THE CELL CYCLE. POSSESSES DNA-
CC BINDING ACTIVITY. IS A NEGATIVE REGULATOR OF THE TATA BOX-BINDING
CC ACTIVITY OF TBP.
CC -!- SUBUNIT: TFIID IS COMPOSED OF TBP AND A VARIETY OF TBP-ASSOCIATED
CC FACTORS.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC -!- SIMILARITY: CONTAINS 1 HMG BOX.
CC -!- SIMILARITY: TO HUMAN TAFII-250 (CCG1). SOME TO S.POMBE TAFII-111
CC AND TO S.CEREVISIAE TAF145.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

```



```

Align seg 1/1 to: P300_HUMAN from: 1 to: 2414

17 CCGCCGCGC.....GCGCGCGCGCCCTGCTGCGCGCGG 54
   |||||
870 PROProGluProGlnSerGlnAlaLeuHisProPro..... 882
55 GGTCTCGGGGGCCCGCTCCGCGCTCGGCGCGGACCGGGA 104
   |||||
883 ...ArgGlnThrProThrProThrThrGlnLeuProGlnGlnValG 898
105 GCGGCGCGCGACGCGCTGGCGCTGGCGGGGGCGGCGACCGGCGCG 154
   |||||
898 InProSerLeuProAlaAlaProSerAlaAspGlnProGlnGlnInPro 914
155 GTCGAGCATGGGCGAAGACCAAGACCAAGCAAGTC..... 190
   |||||
915 ...ArgSerGlnGlnSerThrAlaAlaSerValProThrProAsnAl 929
191 .....GGACAAACACCTCTACGA.....GGAGTAT 215
   |||||
929 aProLeuLeuProProGlnProAlaThrProLeuSerGlnProAlaValS 946
216 GTAGAGAGACCCCTTGACGTGCTCTCAAGTAGAGAGAGAGCAAGTCAC 265
   |||||
946 eTlLeGlnGlnGlnValSerAsnProProSer.....ThrsSer 959
266 CGAAGCTCTCAGCGGCGCTCGGCGCGACGCTCGAGCTCTTGAAAGACA 315
   |||||
960 ThrGlnValAsnSerGlnAlaAlaGlnGlnGlnPro..... 972
316 AAAGCATCATGACAAACAAAGACAGACAGCAAGGAAAAAGAAAGANA 365
   |||||
973 .....SerGlnGlnValSerGlnValSerGlnValSerGlnVal 984
366 GCGAGAGACAGATTCACGAGGAGAAAGGAGAAACGAGAAAGAGT 415
   |||||
984 aLaspGlnProGlnProAlaAspThrGlnProGlnAspLysSerGlnU 1000
416 TAAGGAGATTAATAAGACGAGATCGAGACCGGATGGAATGAGCAG 465
   |||||
1001 LysValGlnAspCysLysMetGlnSerThrGlnThrGln...GlnVal 1016
466 AAAAGATTCACAGTGTACAGCCCTGTGAGATTAGACTTGCCTCTGAG 515
   |||||
1016 eThrGlnLeuLeuSerThrGlnLeuGlnGlnGlnGlnGlnGlnGln 1032
516 AACCTCTCAACAGCTCTTACG.....AAACAGAAAGAGT 553
   |||||
1033 SerAlaThrGlnSerSerProAlaProGlnGlnSerLysLysLysLys 1049
554 AGAAGACAGACCCCTTCAGAACCTTTGAATCAACTGATGAGACAATGC 603
   |||||
1049 eLysProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1066
604 AGAGAAAGATCCAAAGTCTTC...TTTTCATTCTGACTGACTGATT 650
   |||||
1066 yTrArgGlnAspProGlnSerLeuProGlnProGlnProValAspProGln 1082
651 ATT.....GCTCTGCTACTCTCATGATCATTAACACCAATGGATT 694
   |||||
1083 LeuLeuGlnGlnLeuProAspLysPheAspLysLysSerProMetLys 1099
1099 uSerThrLysLysArgLysLeuAspThrGlnGlnGlnGlnGlnGln 1116
695 TACTACATGAAGAAAGAAAGATCAAGACAATGACTATCAGTCCATGAAG 744
   |||||
745 AACTAAAGGTAACCTCAACTATGTGTAATGCCATGATTACAT 794
   |||||
1116 InTyValAspAspLysLysLeuMetPheAsnAsnAlaTrpLeuTyAsn 1132
795 AAACAGAGCCATTTATTAAGGCAAGACAGCTGTTGCACACAG 844
   |||||
1133 ArgLysThrSerArgValTyLysTyLysSerLysLeu..... 1145
845 AATGAATTTCTAGCCAGAAAGAAATTCAGAGCTTACAGACATAG 894
   |||||
1146 .SerGlnValPheGlnGlnGln...LysProValMetGlnSerLeuG 1161
895 ACTTCATGCT..... 905
1161 LyTyCysCysGlyArgLysLeuGlnPheSerProGlnThrLeuCys 1177
906 ...GACTTCAGAAACTCGAAAGCAGAAAGATGAGAACACACACCTCACA 952
1178 TyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1194
953 GAGTGGGAGAGAGAGAGCTGC.....TGCGAGAG 984
   |||||
1194 rGlnAsnArgTyThrHisPheCysGlnLysCysPheAsnGlnLysGln 1211
985 AGAGAGAGAGACTCGGAGATGCCAGACACAGCGCTTCAGAGTCCAGC 1034
   |||||
1211 uSerValSerLeuGlnAspAspProSerGlnProGlnThrThrLys 1227
1035 AAAGAAAT.....AAAGAAAGACAAAGATGCTGAAAGTAAATT.. 1079
   |||||
1228 LysGlnGlnPheSerLysArgLysAsnAspThrLeuAspProGlnLeu 1244
1080 ...AAAGCAATTAATTAGAGAGAGAGAGAGAGAGCTT..... 1115
1244 eValGlnCysThrGlnCysGlyArgLysMetHisGlnLeuGlnVal 1261
1116 .....GACCGCATGCTGAG 1130
1261 lHisGlnLysLysThrProAlaGlnPheValCysAspGlyCysLeu 1277
1131 GAATCTGA..... 1139
1278 LysSerAlaArgThrArgLysGlnAsnLysPheSerAlaLysArgLeu 1294
1140 .....GGAAGCTGACAGCGGCTGTGAAACGACGAGTCG 1176
1294 oSerThrArgLeuGlnGlnThrPheLeuGlnAsnArgLysAsnAspPhe 1311
1177 AATTGAAAGAAAGAAACACAGATGAGACAGAGCTTGGGACTTCCAT 1226
   |||||
1311 rGArgGlnAsnHisProGlnSerGlnGlnValThrValArgValHis 1327
1227 CCTGTGATCCATTTGAGAGAGAGCCAGGCTACGCTGAGAGACTGG 1276
   |||||
1328 AlaSerAspLysThrValGln.....ValLysProGln 1338
1277 AATGACAACTGAGACACTGAGTGAAGTGAATCTTGCAGGGGTCA 1326
   |||||
1338 yMetLysAlaArgPheValAspSerGly...GlnMetAlaGlnSerPhe 1353
1327 AAGAGATTAAGAAAGAAAGATCACTCCAGTGTATATTGAATTATG 1376
1353 ..... 1353
1377 CCTACAGTTCTTATGACCCGATTATGACTCCATTTGCAATATAG 1426
   |||||
1354 ProTyArgThrLysAla.....LeuPheAlaPheGlnGln 1365
1427 CAAGATGATCTGATTATATCTAT.....TCACCTATG 1461
   |||||
1365 uLysAspGlnValAspLysCysPhePheGlnMetHisValGlnGln 1382
1462 GGAAGACTGTGATCTCCAACTGAT..... 1487
   |||||
1382 LysSerAspCysProProAlaGlnArgArgValTyLysSerTyLeu 1398
1488 TTCAGCATCAGATTTTGGCCACAGTGC.....CAAGATTATCC 1528
   |||||
1399 AspSerValHisPhePheArgProLysCysLeuArgThrAlaValTy 1415

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@sb-sib.ch).
CC -----
CC
DR EMBL; X72889; CA51407.1; -.
DR EMBL; D26155; BAA05142.1; -.
DR MIM; 600014; -.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; SNF2_N.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR PRINTS; PRO0503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
KW Transcription regulation; Nuclear protein; Activator; Bromodomain;
KW ATP-binding; Helicase; Alternative splicing.
FT DOMAIN 216 238 POLY-GLN.
FT DOMAIN 241 249 POLY-GLN.
FT DOMAIN 555 558 POLY-ARG.
FT DOMAIN 639 646 POLY-GLU.
FT NP_BIND 745 752 ATP (POTENTIAL).
FT SITE 847 850 DEGH BOX.
FT DOMAIN 1293 1297 POLY-GLU.
FT DOMAIN 1415 1485 POLY-GLU.
FT DOMAIN 1514 1525 BROMODOMAIN.
FT VAAPSLIC 1397 1414 MISSING (IN SHORT ISOFORM).
FT CONFLICT 239 239 P -> PQQP (IN REF. 2).
FT CONFLICT 390 390 Q -> E (IN REF. 2).
FT CONFLICT 509 509 G -> S (IN REF. 2).
FT CONFLICT 707 707 W -> R (IN REF. 2).
FT CONFLICT 1135 1135 D -> H (IN REF. 2).
FT CONFLICT 1390 1390 C -> V (IN REF. 2).
SQ SEQUENCE 1586 AA; 180762 MW; FA537FEA2392807A CRC64;

alignment_scores:
    quality: 210.00      length: 363
    ratio: 1.235         gaps: 15
    percent similarity: 46.832      percent identity: 23.967

alignment_block:
US-09-687-230-1 x SN22_HUMAN ..
Align seg 1/1 to: SN22_HUMAN from: 1 to: 1586

321 GATCATGACAAACACAAAGAC.....AGAAACGGAAAAAGAG 358
    ||| |||:|||||:||||| |||
1253 AspmetaSprgrgrgrgluaspalaArgrAanProlYArgrLsprorAr 1269
    359 AAAGAAAGAGAGAAAGCAGATTCCAGG..... 386
    | ||| |||:|||||:
1269 gremetgluGluaSpgluLeuProserTrpIleIlelyAspaSpalag 1286
    387 .....GAAGAAAG..... 395
1286 luValgluArgrLeuThrCysgluGlulgluGlulgluysIlePheglyArgr 1302
    396 .....GGGAGAAACGAGAG..... 410
1303 GlyserArggluArgrgrAspaValAspIYrserAspaIaleuThrclulY 1319
    410 ..... 410

```

```

1319 seqIntPleuValaIleIleAspIleValLeuGluGluMetGluGluG 1336
411 .....AGAGTTAAGAGATGATGAAAGACCGA.....GATCGAGACCG 449
      |||||.....|||.....|||.....|||.....|||.....|||
1336 luValaIleuValuValuValuValuValuValuValuValuValu 1352
450 GTGGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 499
      |||||.....|||.....|||.....|||.....|||.....|||
1353 AlAluValuValuValuValuValuValuValuValuValuValu 1364
500 AACACTGCTGCTGAGAGGCTGCTGAGAGGCTGCTGAGAGGCTG 540
      |||||.....|||.....|||.....|||.....|||.....|||
1364 gclValuValuValuValuValuValuValuValuValuValu 1381
541 AACAGAGA..... 548
      |||||
1381 yglMetValuValuValuValuValuValuValuValuValuValu 1397
549 GAAGTGAAGACAGACACCCCTTCAAGAGCTTG.....AATCA 586
      |||||.....|||.....|||.....|||.....|||.....|||
1398 AsnValuValuValuValuValuValuValuValuValuValu 1414
587 AGTGAATGAGCAATGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 636
      |||||.....|||.....|||.....|||.....|||.....|||
1414 tserGluValuValuValuValuValuValuValuValuValu 1427
637 CTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 686
      |||||.....|||.....|||.....|||.....|||.....|||
1427 roserValuValuValuValuValuValuValuValuValuValu 1442
687 ATGATTTTATGATGATGATGATGATGATGATGATGATGATGAT 736
      |||||.....|||.....|||.....|||.....|||.....|||
1443 ValAspValuValuValuValuValuValuValuValuValu 1459
737 CATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 786
      |||||.....|||.....|||.....|||.....|||.....|||
1459 tleuGluValuValuValuValuValuValuValuValuValu 1476
787 TTTACATTAACACAGACATTTATTAATTAATTAATTAATTAAT 836
      |||||.....|||.....|||.....|||.....|||.....|||
1476 hrPhaValuValuValuValuValuValuValuValuValuValu 1491
837 CACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 886
      |||||.....|||.....|||.....|||.....|||.....|||
1492 .....G 1492
887 GAGCATGAGCTTATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 936
      |||||.....|||.....|||.....|||.....|||.....|||
1492 nserValuValuValuValuValuValuValuValuValuValu 1508
937 GAACAGACACTCAGACAGCTGAGGAGGAGGAGGAGGAGGAGGAG 986
      |||||.....|||.....|||.....|||.....|||.....|||
1508 ergLValuValuValuValuValuValuValuValuValuValu 1519
987 AGAGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTG 1036
      |||||.....|||.....|||.....|||.....|||.....|||
1520 AspGluValuValuValuValuValuValuValuValuValuValu 1536
1037 AGAAATATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1085
      |||||.....|||.....|||.....|||.....|||.....|||
1536 sleuValuValuValuValuValuValuValuValuValuValu 1553
1086 .....AATTAATTAAGAGGAGGAG 1103
      |||||.....|||.....|||.....|||.....|||.....|||
1553 rgProValuValuValuValuValuValuValuValuValuValu 1569
1104 CAGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAG 1142
      |||||.....|||.....|||.....|||.....|||.....|||
1570 GluGluValuValuValuValuValuValuValuValuValuValu 1582
seq_name: SwissProt_40:YC40_HUMAN

```

```

seq_documentation_block:
ID YC40_HUMAN STANDARD; PRT; 733 AA.
AC Q9ULI0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein KIAA1240 (Fragment).
GN KIAA1240.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 6:337-345(1999).
CC -! SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AB033066; BAA86554.1; -
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; Bromodomain.1.
DR PRINTS; PR00503; Bromodomain.
DR SMART; SM00287; BROMO.1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS00614; BROMODOMAIN_2; 1.
KW Bromodomain; Hypothetical protein.
FT NON_TER 1
FT DOMAIN 1 320 BROMODOMAIN.
FT SEQUENCE 733 AA; 83861 MW; 445D3D109D7817A CRC64;
SO SEQUENCE

alignment_scores:
Quality: 207.50 Length: 553
Ratio: 0.744 Gaps: 25
Percent Similarity: 50.452 Percent Identity: 21.519

alignment_block:
US-09-687-230-1 x YC40_HUMAN ..
Align seg 1/1 to: YC40_HUMAN from: 1 to: 733

405 CGGAGAGAGGTTAAGGAGATGATGAAAGAGGAGGAGGAGGAGGAG 454
      |||||.....|||.....|||.....|||.....|||.....|||
176 GlnArgProIleGluGluValuValuValuValuValuValuValu 192
455 GAATGAGGAGGAGAA.....AAGATCTCCAGTGTCACCGCC..... 488
      |||||.....|||.....|||.....|||.....|||.....|||
192 unsngAlaIserMetAlaProProArgArgArgArgArgArgArgArg 209
489 .....CCTGATGATTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 524
      |||||.....|||.....|||.....|||.....|||.....|||
209 lAmetGluValuValuValuValuValuValuValuValuValuValu 225
525 ACAGAGCTTTTGAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 574
      |||||.....|||.....|||.....|||.....|||.....|||
226 GluSerGluValuValuValuValuValuValuValuValuValuValu 242
575 AGCT.....TTGAATCACTGATGAGCAATTTGCGAGAGAGAGAGATC 615
      |||||.....|||.....|||.....|||.....|||.....|||

```

```

242 uLeuArgLeuPheLeuArgAspValThrLysArgLeuAlaThrAspLysA 259
616 CAAGTCTTTCTTTTCATTTCTCTGTGACTGATTTTATTTGCTCTGCTAC 665
    ||||| ||||| ||||| |||||
259 rGpheaSnIlePheSerLysProValSerAspTyrLeuGlu..... 272
666 TCCATGATCATTAACACCCCAATGGATTTTGTACCATGAAAGAAAGAT 715
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
273 .....ValIleLysGluProMetAspLeuSerThrValIleThrLysI 287
716 CMAAGCAATGACTATCATGCTCATAGAAAGACTAAAGATTAATCTTAAC 765
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
287 eaSPryLHisnIlePheThrLeuThrAlaLysAspPheLeuLysPheAsp 304
766 TAATGTGCTAATATGCGATGATTTACAAATAAACAGAGACCATTAATAT 815
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
304 euIleCysSerAsnAlaLeuGluTyrAsnProAspLys..... 316
816 AAAGCTGCAAGAGCTGTTCACACTCAGCAATGAATAATTTCTTACCCAGA 865
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
317 .....AspProGlyAspLysIleIle.....ArgH1 325
866 AAGAATTCAGACCTGAAGCAGACAGATA..... 893
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
325 saRgaIaCysThrLeuLysAspThrAlaHisAlaIleIleAlaIaGlu 342
894 .....GACTTCATGGCTGACTTCAGAAAGCTCAGAAAGCAGAAAGAT 935
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
342 euAspProGluPheAsnLys..LeuCysGluGluIleLysGluAlaArg 358
936 GGAACAG.....ACACCTCACAGACTCG 958
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
358 euLysArgGlyLeuSerValThrSerGluGluIleAsnProHisSerThr 375
959 GG.....AGCAGCAGGCTGCTGCGCAGAGAGAGAGAG 990
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
375 LysAlaArgLysThrGluThrArgValGluIuAlaPheArgHisLysGlu 391
991 AGGACTGTGGAGATCCGGAAGCAGACCCCTTCAAGAGTCCACAGAAAGAA 1040
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
392 ArgAsnProMetAspValThrHisAsnSerAlaAsnLysCysAlaPheAr 408
1041 AATAAAGAAAGAC.....AAGATATGCTTGA 1069
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
408 gValArgArgLysSerArgArgArgSerGluThrGlyLysGlyIleIle 425
1070 AGATTAAGTTAAAGCAATAATTAGAGAGAGACAGACAGACGCTT.... 1115
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
425 ys..LysArgGlyValAsnAsnLeuLysLysAspGluGluAspThrLysP 441
1116 .....GACCGCATCGTGAAGGAATCTGGA 1139
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
441 heaIaAspTyrGluAsnHisThrGluAspArgLysLeuGluAsnGly 457
1140 GGAAGCTGACACAGCGCGCTTGAACAGTCAGTCAGGATTTGAAAGAG 1189
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
458 .....GluPheGluValSerThrAspCysHisGluGluAsnGly 470
1190 AAAACAGATGGAACAAGACGTTGAGACTTCCATCTGTGATCCA 1239
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
470 yGluGluThrGlyAspLeuSerMet..... 478
1240 TTGTAGAGAGACCGGCTACTGCTGTGAGACTGGAGATGACAACTGGA 1289
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
479 ...ThrAsnAspGluSerSerCysAspIle...MetAspLeuAspGluGly 493
1290 ...AGACTTCAGTCTGAGTGAATACTTTCAGGCGTTCAAAGAGATAA 1336
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
494 GluArgLeuAsnAsnGlyAlaGlyThrLysGluAsnPheAlaSerThrGly 510
1337 AAGGACAAAGTCACTCCAGCTGTATATTTGAATTATGCGCCCTACAGTT 1386
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
510 uGluGluSerSerAsnLysLeuLeuValAsnSerSer.....Sers 525

```

```

1387 CTATATGCACCGCATATTATGACTCCACATTTGCAAAATATACAGAAAGATGAT 1436
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
525 euLeuAsnProGluGluThrSerArg..... 533
1437 TCTGATTTAATCTATTTCACCTATGGGGAAGACTCTGATCTCCAAAGTGA 1486
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
534 .....LysGluThrPheLeuLysGlyAsnCysLeuAsnGly 546
1487 TTTCAGCATCATGAGTTT.....TTGGCCACGCTGCCAAGATT 1524
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
546 uAlaSerThrAspSerPheGluGlyIleProValLeuGluCysGlu.... 561
1525 ATCCGATATGCATCCAGATATGTTACTGATGATTTTAAACAAAGAGAGCG 1574
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
562 .....AsnGlyLysLeuGluValValSerPheCysAsp 572
1575 CATTCACAGACCTTACACAGATGAGAGATGTCAATGCTGAGATGAGAGC 1624
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
573 SerGlyAspLysCysSerSerGluGluLysIleLeuLeuGluAspGlnSe 589
1625 C.....CATACTAGACACTTGACACAGAAAGAAAGATGAGC 1662
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
589 rLysGluLysProGluThrSerThrGluAsnHisGlyAspAspLeuGlu 606
1663 AGATTACAGAAAGTAGAGCCACAGCGGCTTGGACTTCACACTCAAGAG 1712
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
606 yLeuGlu.....AlaLeuGluCysSerAsnAsnGlu 616
1713 AGGCTCATACCGCTGAAGACGATTAACAATTTTGGCTTCCAGTTGAGAT 1762
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
617 LysLeu.....GluProGlySerAspValGluVal 626
1763 TTTTGACTCTGAAAGAGCTGAATAATTCCAGAAAGAACTGATGAGACCA 1812
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
626 LysAspAlaGluLeuAspLysGluGlyAlaSerLysValLysLysTyrA 643
1813 CCAAGTTGCTCAGGGAAGCTCCAGGAAGCCCAAGTGAACGTTTGACACCC 1862
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
643 rGlyLysLeuIle.....LeuGluGluAlaLysThrThrSerLeuGluLeu 657
1863 AGACCC 1868
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
658 ValPro 659
seq_name: SwissProt_40:SPT7_YEAST
seq_documentation_block:
ID SPT7_YEAST STANDARD: PRT: 1332 AA.
AC P35177.
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Transcriptional activator SPT7.
GN SPT7 OR YBR081C OR YBR0739.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95229044; PubMed=7713415;
RA Gansheroff L.J., Dollard C., Tan P., Winston F.;
RT "The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein
RT important for transcription in vivo.";
RL Genetics 139:523-536(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95076715; PubMed=7985423;
RA van der Aart O.J.M., Barthe C., Doignon F., Aigle M., Crouzet M.,
RA Steensma H.Y.;

```



```

640 rvaIhISGLuAsnValaIsnLysasnclutleLysnGluasnGlyLysasnG 657
1099 GAGAGCAGAGGAGCAGCTTGACCCGATCGTAGAGGATCTGG.....AGG 1141
657 lueIuGlnAspMetValaIGluLysSerLysrthGlnAspSerSerLys 673
1142 AAGCTGACCGAGCGGCTTGACAGCTGACGTCGCAATTGTGAAGAGAA 1191
674 AspAlaAspAlaAlaLysLysAspThrGluAspGlyLeu...GlnAspLys 689
1192 AACCAAGATGAAACAAGACAGCTTGACCTTCGATCCCTGATGATCCAT 1240
689 srtHrAlaGluAsnLysGluAlaGlyLysAsnGlnGluLysGluLysAsp 706
1241 .....GTAGAGAGCCAGGCTTA 1258
706 spAspAspGluAspGluAspGluAspMetValaLAspSerGlnSerLysLeu 722
1259 CTGCGCTGTCGACCTGGCAATGACAACCTGAGACTTCAGCTGG..... 1303
723 LeuGluLysAspAspAspAspAspAspLeuGluLysSerValTrpLysTh 739
1304 .....AGTGAATACT 1313
739 rvaIThrAlaLysValaArgAlaGluLeuGlySerLysAspThrGlyTrp 756
1314 TT.....GCAGGGGTTCAAGAGGA 1333
756 helLysasnGlyLysLeuAsnSerAspSerGluAlaPheLeu.LysasnPr 772
1334 TAAAGCAACAAGATCAGCTGTTATATTTGAATTATGGCCCCCTACA 1383
772 oclnArgMetLysArgPheAspGlnLeuPheLeuGluTrpLysGluGlnL 789
1384 GTTCTTATGACCCGATTCATGACCTCCACATTTGCCAAT..... 1421
789 ysaLleuGlnLysSerLysrthArgGlnLysIleGluGlnAsnSerIleMetLys 805
1422 .....ATGACAGAGATGATTTGTGTTAATC..... 1448
806 AsnGlyPheGlyThrValLeuLysGlnLysAspAspGlnLeuGlnPhe 822
1449 .....TATTCACCTATGGGAGAGACTCT.....GATCTTCCA 1482
822 ehlAsnAspHisSerLeuAsnGlyAsnGluAlaPheGlyLysGlnProA 839
1483 GTGATTTTACGATCCATGAG.....TTTTGGCCACGTCGCCAAGATTAT 1536
839 snAspIleGlnLeuAspAspThrArgPheLeu.....GlnGluTrp 852
1537 CGGTATGTCATGGCAGATAGTTTACTGATGTTTAAACAAAGAGGCA 1576
853 Asp.....IleSerAsnAlaIleProAspIleValGlyGluGlyValas 867
1577 TTCAGAGACCTTACAGAGATGAGATCTCATTCCTGAGATGAGAGGCC 1626
867 nThrLysThrLeuAspLysMet.....GlnAspLysSer 879
1627 ATACAGAGACACTTGACACAGAGAAAGAAATGAGAGATTAACAGAGTA 1676
879 alaAspArgMetLeuGlnAsnGly..... 886
1677 GAGCCAGCAGGCGCTTGACTCCAGTACTCAAGACAGGCTCATAGGCT 1726
887 .....IleAsnLysGlnSerArgPheLeuAlaAs 896
1727 GAAAGCACTAACAAATTTGGCGTTCCAGTTGAAGTTTGGACTCTGAG 1776
896 nLys.....AspLeuGlyLeuTrpProLysMet.....AsnGlnA 908
1777 AAGCGAAATATTTCCAG.....AAGAAACGTTGATGACACACAGATTG 1820
:: .....::

```

```

908 snlLeuThrLeuIleGlnGlnIleArgHisIleCysHisLysIleSerLeu 924
1821 CTCAGGGAACCTCCAG.....GAGCCCAAGATGACCT 1853
925 IleArgMetLeuGlnSerProLeuSerValaGlnAsnSerArg 938
seq_name: SwissProt_40: ATRX_CAEEL

```

```

seq_documentation_block:
ID ATRX_CAEEL STANDARD; PRT; 1359 AA.
AC G9U7E0; 002061.
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).
GN XNP-1 OR B0041.7
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
RX MEDLINE=99365296; PubMed=10433961;
RA Vallard L., Fontes M., Ewbank J.J.;
RT "Characterization of xnp-1, a Caenorhabditis elegans gene similar to
RL the human XNP/ATR-X gene."
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2.
RA Fulton R., Wohlmann P.;
RL Submitted (Jun-1997) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC GENE EXPRESSION BY AFFECTING CHROMATIN (Potential).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
DR EMBL; AF134186; AAD53361.1; -
DR EMBL; AF000196; AAC24256.1; -
DR WormPeP; B0041.7; CE17314.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00271; hellicase_C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
KW DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding.
FT NP_BIND 496 503
FT SITE 636 639
FT DOMAIN 67 70
FT DOMAIN 266 272
FT DOMAIN 372 375
FT DOMAIN 603 608
FT DOMAIN 859 862
FT CONFLICT 479 479 C -> F (in Ref. 2).
SQ SEQUENCE 1359 AA; 156191 MW; EBA4342547D4F4E64 CRC64;

```

```

alignment_scores:
Quality: 198.50 Length: 660
Ratio: 0.577 Gaps: 23
Percent Similarity: 52.121 Percent Identity: 21.061

```

alignment_block:
US-09-687-230-1 x ATRX_CAEEL ..
Align seg 1/1 to: ATRX_CAEEL from: 1 to: 1359

```
129 GCGCGGGGGGGCGCGACGGGGCGCGTGGGACATGGCGCAAGAACACAA 178  
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
86 AAlaySSerGIuSerGIuSerAspGIuGIuGIuSerAspGIu 102  
||| ||||| : : : : : : : : : : : : : : : : : : : : : : : :  
179 GAAACACAAAGTCGACAAACCTCTACAGAGAGTATGTAGAAAGCCCT 228  
||| ||||| : : : : : : : : : : : : : : : : : : : : : : : :  
102 eLySSerLySSerLySlySlyValAspGIuLySlySlyGIuLySSerL 119  
229 TGAACCTGGTCTCAAGTAGAGAGGAACGACCGCAACTCTCCACG 278  
||| : : : : : : : : : : : : : : : : : : : : : : : : : :  
119 yLyS.....LysArgThrThr 124  
279 GGCACCTCGGGCAGACCTCAGCCTCTCGAAGACAAAGATCATGCA 328  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
125 SerSerSer.....GluAspGIuAspSerAspGIu 134  
329 CAACACAAAGACAGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 378  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
134 uGIuArgGIuGIuLySerLySlySlySerLySlyThrLySlyGIuT 151  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
379 TTCAGAGGGA.....GAAAGGGAGAGGAGGAGGAGGAGGAGGAGG 422  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
151 hSerSerGIuSerGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 167  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
423 GATAAAAAGACGAGAGATCGAGAC.....CGGCTGAGCAATGAGGC 463  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
168 SerLySlyAsnLySlyGIuLySerValLySlyArgAlaGIuThrSerL 184  
464 AGAAAAAGATCTCCAGTGTACGCGCCCTGTGATTAAGTCTGCTCG 513  
||| : : : : : : : : : : : : : : : : : : : : : : : : : :  
184 uGIuSerAsp.....GluAspGIuLySproS 193  
514 AGAAGCTCTCACAGCTCTTACGCCAAACAGAGACAGTACAGACACA 563  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
193 eLyLySerLySlySlyGIuLySlyValLySlySerGIuSerGIu 209  
564 CCCCCTCAAGAGCTTGAATCAACGATGAGACATTCAGAGGAGAGAGA 613  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
210 SerGIuSerGIuAspGIuLySlyGIu...ValLySlySerLySlySlySe 225  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
614 TCCAGTGTCTTCTTTCATTTCTGTGACTGATTTTATTTGCTGCTGCT 663  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
225 rLySlyValLySlyGIuSerGIuSerGIuAspGIuAlaProGIuL 242  
664 ACTCCATGATCATTAACACCCCAATGGATTTTGTACCATGAAGAGAA 713  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
242 yLySlyThrGIuLySlyArgLySlySerLySlyThrSerGIuGIuSer 258  
714 ATCAAGAACATGATCATTCATAGAGAACTTAAGATTAACCTTCA 763  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
259 SerGIuSerGIuLySlySerAspGIuGIuGIuGIuGIuGIuGIuSer 275  
764 ACTAATGTACTAATGCCATGTATTACATAAACCAGAGACATTTATT 813  
||| : : : : : : : : : : : : : : : : : : : : : : : : : :  
275 O.....LysProLySlySlySlyS 282  
814 ATAAGCTGCAAGAGAGCTGTG.....CACTCAGCA 845  
||| : : : : : : : : : : : : : : : : : : : : : : : : : :  
282 rOleuAlaValLySlyLeuSerSerAspGIuGIuGIuGIuGIuSerAsp 298  
846 ATCAAAATTTTACGCCAGGAGAAATTCAGAGCTGAGAGAGACATAGA 895  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
299 ValGIuValLeuProGIuLySlySlyArgGIuAlaValThrLeuIleSe 315  
896 CTTTCATGCTGACTTCGAGAAACTCGAAAGCAGAGAAATGATGAGACACA 945  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
315 rAspSerGIuAspGIuLySlyAspGIuLySlySerGIuSerGIuAlaSerAsp 332
```

```
946 CPTCACAGAGTGGGAGACGAGCGAGCTGTGCGAGAGAGAGAGAGAC 995  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
332 aGIuGIuLySlyValSerLySlySlyAlaLySlyGIuGIuSerGIu 348  
996 TCTGAGATGCCGAGACACACGCTTCAAGATGCCGAGGAGGAGGAGG 1045  
||| : : : : : : : : : : : : : : : : : : : : : : : : : :  
349 SerGIuSerAspSerGIuGIuSerLethrValAsnArgLySlySerLy 365  
1046 AAGAGAGACAAAGATATGCTTGAAGATAG.....TTTAAAGCA 1086  
||| : : : : : : : : : : : : : : : : : : : : : : : : : :  
365 sLySlyGIuLySlyProGIuLySlySlySlyGIuLleIleMetLaspers 382  
1087 ATAAATTAAGACAGACGAGCAGCAGCTTGACCGCATCTGAGAGAACT 1136  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
382 eLySlyGIuGIuGIuThrLleAspAlaGIuArgAlaGIuLySlyGIuArg 398  
1137 GGAGAAAGCTGACAGCGGCTGTGACAGTGCAGTGCAGATTTGAAG 1186  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
399 .....ArgLySlyArgLeuGIuLySlySlyGIuSlySlySlySlySly 409  
1187 AAGAAACCAAGATGGAACMACAGAGCTTGGGACTTCTCCATCTGTGATC 1236  
410 .....AsnG 411  
1237 CCATTGTAGAGACCGCAGCTACTGCGTGGAGACTGGGAAATGCAACT 1286  
||| : : : : : : : : : : : : : : : : : : : : : : : : : :  
411 yLleValLeuGIuGIuLyGIuLySlySlySlySlySlySlySlySly 425  
1287 GGA.....AGACTTCAGTCTGAGTGAATACTTTTGACAGG 1321  
||| : : : : : : : : : : : : : : : : : : : : : : : : : :  
426 GlyThrSerSerGIuArgLySlyLeuLySlySerValLyLeuAspProAsp 442  
1322 GTTCAAGAGGATTAAGAAACAAAGTACCTCCAGTGTATATTGAT 1370  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
442 rSerThrValAspGIuGIuGIuLySlySlySlySlySlySlySlySly 458  
1371 .....TATGGCCCTCAGTCTTATGACCGCAT..... 1400  
458 eLLeuValArgLleLeuLySlyProHISGlnAlaHISGlyLleGlnPheMet 474  
1401 TATGACTCCACATTTGCAAAATTCAGCAAGATGATTTGATTTAATGTA 1450  
||| : : : : : : : : : : : : : : : : : : : : : : : : : :  
475 TyrAspCysAlaCysGIuSerLeuAspArgLeuAspThrGIuGIuSerGI 491  
1451 TTCAACCTATGGGAAAGCTGATCTCCAGATGATTCAGATCCATCATG 1500  
491 yGIuLleuAlaHIScysMetGIuLeuGIuLySlyThrLeuGIuAlaLiet 508  
1501 AGTTTTGGCCAGCTGCCAAGATTAATCCGTATGTC.....ATGGCAGAT 1544  
||| : : : : : : : : : : : : : : : : : : : : : : : : : :  
508 hPheLeuHISThrValLeuMetHISGIuLySlyLleGIuLySlyCysLyS 524  
1545 AGTTTACTGATGATTTTAAACAAAGAGGAGGATTCAGACCCCTCAAGA 1594  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
525 ArgValLeuValValProLySlySlyValLleIleAsnThrPheLySGL 541  
1595 GATGAGATGATGATGCTTGAAGATGAGGAGGATTCAGACACTTGACA 1644  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
541 uPheGIuLySlyThrPheValAspAsnAspGIuGIuLeuAspThrIleAsp 558  
1645 CAGGAAAGAAATGGAGCAGATTACAGAAATGAGACCAACGAGGCTTGG 1694  
: : : : : : : : : : : : : : : : : : : : : : : : : : : :  
558 aLAsnGIu.....Leu 561  
1695 GACTGC.....AGTACTCAGACAGGCTCATAGCGCTGAAGACA..... 1733  
||| : : : : : : : : : : : : : : : : : : : : : : : : : :  
562 AspSerLySlyThrLleGIuSlySlyArgAlaLleLySlyAlaLleTIPH 578  
1734 .....GTAACAATTTGGCGTTCAGTTGAAGTTT 1764  
578 sSerSerLySlyThrProSerValMetLleIleGIuTyrAspLeuPheArgI 595  
1765 TTGACTCTGACAGAGCTGAATATTCCAGAGAAAGAACTTGATGAGCCACC 1814
```


1425AGCAAGATGATTCTGATTTAATCTATTCA 1454


```

678 AACACCCATGATTTAGTACCATGAAGAAAGATCAAGAAACATGA 727
|||||
331 LysHisProMetAspLeuSerThrValIAspArgLysMetGluAsnArgAs 407
728 CTATCAGTCATAGACAACTAAAGGATTAACCTAACTATGCTGTACTA 777
|||||
407 PTyArgAspAlaGlnGlnIlePheAlaAlaAspValArgLeuMetHeserA 424
778 ATGCCATGATTTACATTAACCCAGACCACTTTATTAAGCTGCAAG 827
|||||
424 sncysTYrLYsTYrAsnProProAspHisAspValValAlaMetAlaArg 440
828 AAGCTG..... 833
441 LysLeuGlnAspValPheGluPheArgTYrAlaLysMetProAspGluPr 457
834 .TTGCACCTACGAAATGAAATTTCTTACCCAGAAAGATTCAGAGCTGA 882
|||||
457 oLeuGluProGlyProLeuProValSerThrIleMetProProGlyLeu. 473
883 AGCAGACATTAAGACTTCATGCTGCTGCTGCAAAAACCTCGAAACGAGAA 932
|||||
474 .....AlaLysSerSer 477
933 GATGGAACAGACACCTCAGAGAGTGGGAGGAGGAGCTGCTGGCAGAG 982
|||||
478 SerGluSerSerGluGluSerSerSerGluSerSerGluGluGlu 494
983 AGACAGAGAGGACTCTGAGATGCCGCAAGCACAGCCTTCAGAGTCCCA 1032
|||||
494 uGluGluGluAspGluGluAspGluGlu..... 503
1033 GCAAGAAATTAATAAAGAAAGACAAATATGCTTGAAGATTAA 1082
|||||
504 .....Arg 516
1083 AGCAATATTATTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1132
|||||
517 AlaHisArgLeuAlaGlnIleGlnGlnIleGln.....ArgAlaValHisGlu 532
1133 ATCTGGAGGAAAGCTGACAGG 1154
|||||
532 uGlnLeuAlaAlaLeuSerGln 539
seq_name: SwissProt_40:TRDN_HUMAN
seq_documentation_block:
ID TRDN_HUMAN STANDARD: PRT: 728 AA.
AC Q13061;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Triadin.
GN TRDN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=96061957; PubMed=7588753;
RA Taske N.L., Eyre H.J., O'Brien R.O., Sutherland G.R., Denborough M.A.,
RA Foster P.S.;
RT "Molecular cloning of the cDNA encoding human skeletal muscle triadin
RT and its localisation to chromosome 6q22-6q23.";
RL But. J. Biochem. 233:258-265(1995).
CC -1- EDUCATION: MAY BE INVOLVED IN ANCHORING CALDESQUESTRIN TO THE
CC JUNCTIONAL SARCOPLASMIC RETICULUM AND ALLOWING ITS FUNCTIONAL
CC COUPLING WITH THE RYANODINE RECEPTOR (BY SIMILARITY).
CC SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN, SARCOPLASMIC
CC RETICULUM.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 018985; AAA75315.1; -.
ID MIM: 603283;
KW Transmembrane; Sarcoplasmic reticulum; Glycoprotein.
FT INIT_MET 0 46 BY SIMILARITY.
FT DOMAIN 1 6 CITOPLASMIC.
FT TRANSMEM 47 67 POTENTIAL.
FT DOMAIN 68 728 POTENTIAL.
FT CARBOHYD 74 74 LUMENAL.
FT FT 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 728 AA: 81423 MW; C1C53BBE1B20815 CRC64;

```

alignment_scores:

Quality:	170.00	Length:	656
Ratio:	0.588	Gaps:	28
Percent Similarity:	44.035	Percent Identity:	20.732

alignment_block:

US-09-687-230-1 x TRDN_HUMAN

Align seg 1/1 to: TRDN_HUMAN from: 1 to: 728

```

225 CCCTTGAAGCTGCTCTCAAGTAGAGGAAAGCAAGTCACGCA..... 269
|||||
87 ProLeuLysLeuVal...ArgAspAlaMetGluGluThrThrAspTrpI 102
270 .....CTCTCCACGGGACACTCGGGGCGACGACTCCA 300
|||||
102 eTYrGlyPhePheSerLeuLeuSerAspIleIleSerSerGluAspGlu 119
301 GCCTTCTGAGACAAACAAAGCATGACACAA..... 332
|||||
119 LysAspAspAspGlyAspGluAspThrAspIleGluIleAspGluPro 135
333 .....CACAAAGACAGAAACGGGAAAGAGAA 361
|||||
136 ProLeuArgLysLysGluIleHisLysAsp...LysThrGluLysGlnI 151
362 GAAAGGAGAGAGAGAGAT.....CCAGGGAAGAA 393
|||||
151 LysProGluArgLysIleGlnThrLysValThrHisLysGluLysGlu 168
394 AGCGGAGAAACGAGAAAGAGTTAAGAG.....GATTAAGAAAGAGCA 437
|||||
168 ySGlyLysGluLysValArgGluLysGluLysProGluLysLysAlaThr 184
438 GATCGAGACCGGGTGAGAAATGAGCGCAAAAGATCTCAAGTGCAGCG 487
|||||
185 HisLysGluLysIleGluLysLys..... 192
488 CCCTGTAGATTAGACTTGCTCTGAGAGAGCTCTCAACAGCTCTTAG 537
|||||
193 .....GluLysProGluThrLysTrpValA 201
538 CCAACACAGAA..... 548
|||||
201 lalysGluGlnLysLysAlaLysThrAlaGluLysSerGluLysThr 217
548 ..... 548
218 LysLysGluValLysGlyLysGlnGluLysValLysGlnThrAlaAl 234
549 .....GAAGTAAGAACAGACACCC.....CTTCAAG 573
|||||
234 alysValLysGluValGlnLysThrProSerLysProLysGluLysGluA 251

```

```

574 AACCTTGAATCACTGATGAGACAAATTCAGAGAAAGATCCAGTGC 623
    ::      ::      ::      ::      ::      ::      |||
251 sPlYsGluLysAlaIaValSerLysHISGluInLysAspInLysAla 267
    |||
624 TTCTTTTCAATTCCTGTGCTGCTGATTTTAT 661
    |||      ::      ::      ::      ::      ::      |||
268 pHeCysAlaGTYrMetIleAspIlePheValHISGluAspLeuLysProG 284
    |||
662 CTACTCCATGATCATTAAACACCCCAATGCAATTTGCT 698
    |      |||      |||      ::      ::      ::
284 yGInSerProAlaIleProProIleuProThrGluInAlaSerArgP 301
699 ..... ACCATGAAAGAAAGATCAAGAACAAATGACTAT 731
301 rOThrProAlaSerProAlaLeuGluLys 311
732 CAGTCCATGAGAAAGAAATTAAGTCAATTCCTCAATGCTACTAATGC 781
    |||      |||      ::      ::      |||
312 ..... GluGluLysLysAlaGluLysValThrSerG 325
782 CATGATTTTCAATTAACAGACATTTATTTAAAGCTGCAAGAAC 831
    ::      ::      |||      |||      ::      ::
325 uThrLysLysGluLysGluAspIleLysLysSerGluLysGlu 341
832 TGTTCACACAGAAATGAATTCCTTACCCAGAAAGAAATTCAGACCTG 881
    ::      ::      |||      ::      ::      |||
342 ThraIleAspValGluLysLysGluProGluLysAlaSerGlu Thr 357
882 AAGCAG ..... AGCATGACCTTCATGCTGCTGACTTGCAGAAAAC 922
    |||      ::      |||      ::      |||      ::
358 LysGInGluYThrValLysIleAlaIaGluInAlaIaLysAsp 374
923 AAGCAGAAAGATGGAACAGACACCTCACAGATGGGAG 962
    ::      ::      |||      ::      ::      |||
374 uLysLysGluAspSerLysThrLysLysProAlaGluValGluInP 391
963 .. GACGAGGCTGCTGGCAGAGAGAGAGAGACTCTGGAATGCCGAA 1010
    |||      ::      ::      ::      ::      ::
391 rOlySgLy ..... LysLysGluInLysLysGluLysHIS 402
1011 GCACACGCTTCAGAGTCCCAAGAAAGAAATTAAGAAAGAACCAAA 1060
    ::      |||      |||      ::      ::      ::
403 ValGluProAlaLysSerPro ..... LysLysGluInHIS 414
1061 TATGCTTGAAGATACTTTAAAGCAATTAATTAGAGAGAGACAGAC 1110
    ::      ::      |||      ::      ::      |||
414 rValProSerAspLysGluValLysAlaLysThrGluArgAlaLysGlu 431
1111 AGCTTACCGCATGTGAAGGAATCTGA ..... GGAAGCTGACCAG 1154
    ::      ::      ::      ::      ::      |||
431 IuIleGluAlaValSerSerLysLysAlaValProGluLysLysGluIn 447
1155 CGGCTTGTAAGCATGTCAGTCCGAATTTGAGAAAGAAACCATGAGAAC 1204
    ::      ::      ::      ::      |||      |||
448 LysThrThrLysThr... ValGluInGluInIleArgLysGluLysSerG 463
1205 AAGCAGTTTGGAATCTTCATCCTGTCGATCCATTTAGAGAGACCAG 1254
    |||      ::      ::      ::      |||      ::
463 yLysThrSerSerIleLeuLysAspLysGluProIleLysGlyLysGlu 480
1255 GCTACTGCTGCTG ..... 1268
480 LuLysValProAlaSerLeuLysGluLysGluProGluThrLysLysAsp 496
1269 ... AGACTGGGAATGACACTGGAAGACTCAGTGGAGTGAATACTTT 1315
497 GluLysMetSerLysAlaGlyLysGluValLysProLysProProGluIn 513
1316 GCAGGGGTTCAAGAGAGATAA ..... AGAACAAAGTCA 1350
    |||      |||      ::      ::      ::      ::
513 uGInGlyLysGluLysGluLysProGluProGluInIleLysGluAlaL 530

```

```

1351 CTCACGTGTATATTTGAATTAAGGCCCTACAGTCTTATGACCGCAT 1400
    |||      ::
530 ySProAlaIle ..... 533
1401 TATAGCTCCATTTGCAAAATATGACAGATGATTCGATTTAATCTA 1450
    |||      ::      ::      ::      ::      ::
534 ..... SerGluLysValGluIleLys ..... GluAspIleVal 546
1451 TTCACCTATGGGAGACCTGATCTTCAGTGAATGATTCAGATCCATG 1500
    ::      ::      |||      ::      ::      ::
546 sProGluLysThrValSerHISGluLysProGluGluLysValLeuLysG 563
1501 AGTTTGGCC ..... ACGTGCCAAGATTATCCGTATGTC 1535
    ::      |||      |||      ::      ::      |||
563 InValLysAlaValThrIleGluLysThrAlaLysProLysPro ..... 577
1536 ATGGCAGTATTTACTGATGTTTAAACAAAGAGGACATTCACAGAC 1585
    |||      |||      ::      ::      |||
578 ..... ThrLysLysAlaGluHISArg 585
1586 CCTACAA ..... GAGATGAGATGTCATTCCTGAGATGAG 1623
    ::      ::      ::      ::      |||
585 uArgGluProProSerIleLysThrAspLysProLysProThrProLysG 602
1624 GCCATCTAGACACTTACACAGAGAAAGAAATGAGACCATTAACGAA 1673
    |||      ::      ::      ::      ::      |||
602 yLysThrSerGluValThrGluSerGluLysLysThrGluIleSerGlu 618
1674 GNAGAGCCACAGGCGTTTGACTCC ..... 1700
    |||      ::      |||
619 LysGluSerLysGluLysAlaAspMetLysHISLeuArgGluLysVal 635
1701 .AGTACTCAAGACAGCTCATAGCCCTGAAAGCAATTAATTTGGCG 1749
    |||      ::      ::      ::      |||      ::
635 lSerThrArgLysGluSerLeuGluLeuHISAsnValThrLysAlaGlu 652
1750 TTCCAGTGAAGTTTTCAGCTGTGAAGAGCTGAATATTCAGAAAGAA 1799
    |||      ::      ::      ::      |||      ::
652 ySProAlaArgValSerLysAspValGluAspValProAlaSerLysLys 668
1800 CTTGATGAGACCACC 1814
    |||      |||
669 AlalysGluGlyThr 673

```

Sequence	Strd	Orig	Score	Len	Documentation
sp_human:09NP11	+	3293.50	4422.16	5.9e-239	651 09np11 homo sapiens (human). b
sp_human:09UH55	+	3287.00	4416.38	1.8e-238	652 09uh55 homo sapiens (human). b
sp_rhodet:08B655	+	3262.50	3978.96	4.2e-214	651 08b655 mus musculus (mouse). s
sp_human:09BV48	+	2283.00	3066.27	4.1e-163	451 09bv48 homo sapiens (human). b
sp_human:096K44	+	1716.50	2305.14	1.3e-120	351 096k44 homo sapiens (human). cc
sp_invertebrate:09VLX2	+	738.50	978.29	4.4e-47	861 09v1x2 drosophila melanogaster
sp_human:09NH82	+	684.00	909.94	4.8e-43	501 09nh82 homo sapiens (human). cc
sp_rhodet:09C7M2	+	635.50	854.95	1.7e-39	174 09c7m2 mus musculus (mouse). l
sp_invertebrate:0175A1	+	517.50	683.24	1.6e-30	636 0175a1 caenorhabditis elegans
sp_human:09H789	+	391.00	552.20	4.1e-21	233 09h789 homo sapiens (human). cc
sp_human:09H505	+	274.00	360.31	2.8e-12	362 09h505 homo sapiens (human). cc
sp_human:09Y403	+	255.50	338.94	7.8e-11	715 09y403 homo sapiens (human). cc
sp_human:09UN01	+	252.00	348.29	8.3e-11	56 09un01 homo sapiens (human). ft
sp_plant:09SF11	+	234.50	311.41	5.5e-10	952 09sf11 oryza sativa (rice). sim
sp_rhodet:09SVV4	+	236.00	305.03	2.2e-09	556 09sfv1 arabidopsis thaliana (m
sp_plant:09LUM2	+	229.00	294.09	7.5e-09	652 09plv4 mus musculus (mouse). mc
sp_human:09S682	+	229.00	292.10	7.8e-09	805 09s682 arabidopsis thaliana (mc
sp_invertebrate:09V4J4	+	224.00	279.93	2.1e-08	1430 09v4j4 drosophila melanogaster
sp_invertebrate:097068	+	217.50	267.70	7.0e-08	2065 097068 drosophila melanogaster
sp_vertebrate:09Q941+	+	207.50	236.44	3.7e-07	1633 09q941 gallus gallus (chicken)
sp_invertebrate:076561	+	204.00	265.64	5.0e-07	1235 076561 caenorhabditis elegans
sp_human:09HD59	+	202.00	251.66	9.1e-07	1235 09hd59 homo sapiens (human). b
sp_human:09NH31	+	201.00	247.97	1.1e-06	1582 09nh31 homo sapiens (human). b
sp_human:015355	+	198.00	250.89	1.6e-06	757 015355 homo sapiens (human). mc
sp_plant:09F1A2	+	197.00	265.16	1.4e-06	145 09f1a2 arabidopsis thaliana (m
sp_human:09H273	+	196.50	241.91	2.3e-06	1582 09h273 homo sapiens (human). b
sp_human:09H255	+	196.50	241.79	2.5e-06	1602 09h255 homo sapiens (human). b
sp_human:09H314	+	196.50	241.60	2.5e-06	1634 09h314 homo sapiens (human). b
sp_invertebrate:09V4Y3	+	196.50	235.61	2.9e-06	3080 09v4y3 drosophila melanogaster
sp_vertebrate:09O753	+	195.00	239.60	3.2e-06	1630 09o753 gallus gallus (chicken)
sp_human:09NH03	+	194.00	238.16	3.2e-06	1647 09nh03 homo sapiens (human). s
sp_vertebrate:09O755	+	193.50	237.95	4.2e-06	1568 09o755 gallus gallus (chicken)
sp_vertebrate:09O971	+	193.00	244.51	3.9e-06	723 09o971 gallus gallus (chicken)
sp_vertebrate:073897	+	193.00	244.46	3.9e-06	733 073897 gallus gallus (chicken)
sp_human:09NH04	+	193.00	236.63	4.6e-06	1679 09nh04 homo sapiens (human). s
sp_human:09B8X3	+	193.00	236.44	5.3e-06	1484 09b8x3 homo sapiens (human). s
sp_fungi:09Y7U0	+	191.50	244.68	4.8e-06	578 09y7u0 schizosaccharomyces pom
sp_fungi:043178	+	191.50	240.29	5.3e-06	920 043178 homo sapiens (human). th
sp_human:096396	+	191.50	240.29	5.3e-06	920 096396 homo sapiens (human). th

[illegible]

84 rgvallysgluaspplysllysarqasparqvalgluasnglu 100

462 GCAGAAAAAGATCTCCAGTGTACGCCCCGTGAGATTAGACTTGCCCTCC 511
 |||||
 101 AlAGluLysAspLeuGlnCysHisAlaProValArgLeuAspLeuProPr 117
 |||||
 512 TGAGAAAGCCTCTCACAGCTCTTTAGCCAAACAGAAAGATGAGAACGA 561
 |||||
 117 oGluLysProLeuThrSerSerLeuAlaLysGlnGluValGluGlnT 134
 |||||
 562 CACCCCTTCAGAGACTTTGAATCAACTGATGAGACAAATTGCAGAGAAA 611
 |||||
 134 hPLeuGlnGlnLysAlaLeuAsnGlnLeuMetArgGlnLeuGlnArgLys 150
 |||||
 612 GATCCAAAGTCTTTCTTTTCATTTCCCTGTGACTGATTTTATTCCTCCG 661
 |||||
 151 AsPProSerAlaPhePheSerPheProValThrAspPheLeuAlaProGl 167
 |||||
 662 CTACCTCCAGATCATTTAAACCCCAATGGATTGTACCATGAAGAAA 711
 |||||
 167 yTrSerMetLeuLeuLysHisProMetAspPheSerThrMetLysGluL 184
 |||||
 712 AGATCAAGAACATGACTATCAGTCCATGAGAACTAAAGATTAATCTTC 761
 |||||
 184 ySLeuLysAsnAsnAspArgLysLeuGlnGluLysAspAsnPhe 200
 |||||
 762 AAATGATGTGACTATATGATCCATGATTTACATTAACCCAGAGACCATTTA 811
 |||||
 201 LysLeuMetCysThrAsnAlaMetLeuLysAsnLysProGluThrLeuTy 217
 |||||
 812 TTTTAAAGCTGCAGAAAGACTGTGTGACTCAGAGATGAAGAAATTTCTTAC 861
 |||||
 217 tTrLysAlaAlaLysLysLeuLeuHisSerGlyMetLysIleLeuSerG 234
 |||||
 862 AGGAAAGAAATTCAGAGCCCTGAGACAGACATGACTTTCATGGCTGCTTG 911
 |||||
 234 InGluArgIleGlnSerLeuLysGlnSerIleAspPheMetAlaAspLeu 250
 |||||
 912 CAGAAAACCTGGAAGCAGAAAGATGAGACAGACCTCAGACAGTGGGGA 961
 |||||
 251 GlnLysThrArgLysGlnLysAspGlyThrAspThrSerGlnSerIleGly 267
 |||||
 962 GGAAGAGAGCTGCTGGCAGAGAGAGAGAGAGACTGTGAGATGCCGAG 1011
 |||||
 267 uAspGlyGlyCysTrpGlnArgLysGlnArgLysAspSerGlyAspAlaGln 284
 |||||
 1012 CACAGCGCTTCAGAGATCCCAAGCAAGAAATTAAGAAAGCAAGAT 1061
 |||||
 284 lAlHisAlaPheLysSerProSerLysGlnAsnLysLysLysAspLysAsp 300
 |||||
 1062 ATGCTTGAAGATTAAGTTTAAAGCAATTAATTAGAGAGAGAGAGAGCA 1111
 |||||
 301 MetLeuGlnLysPheLysSerAsnAsnLeuGlnArgGlnGlnGln 317
 |||||
 1112 GCTTGAAGCAGATCGTGAAGAAATCTGAGAGAAAGCTGACAGCGGCTTG 1161
 |||||
 317 nLeuAspArgIleValLysGlnSerGlyLysLeuThrArgArgLeuVal 334
 |||||
 1162 TGAACATCTGCTGCAATTTGAAGAAAGAAACCATGGAACAGACAGC 1211
 |||||
 334 aLanSerGlnCysGlnPheGlnArgLysProAspGlyThrThrThr 350
 |||||
 1212 TTTGGACTTTCATCCCTGTGATCCCATTTGTAGAGAGAGAGAGCTACTG 1261
 |||||
 351 LeuLysLeuLeuHisProValAspProIleValGlyLysProGlyTrpCys 367
 |||||
 1262 CCTGTGTGAGACTGGGAAATGACACTGGAAGACTTTCAGTCTGGAGTGAATA 1311
 |||||
 367 sProValArgLeuGlnGlyMetThrThrGlyArgLeuGlnSerGlyValAsnT 384
 |||||
 1312 CTTTGGCAGGGGCTCAAGAGAGATTAAGAAAGAAAGTCACTCCCGTGTTA 1361
 |||||
 384 hTrLeuGlnGlyPheLysGlnLysPheLysAsnLysValThrProValLeu 400

1362 TATTGGAATTAATGGCCCTPACAGTTCTTATGACACCGCATTAATGACTCCAC 1411
 |||||
 401 TrLeuLysnTrpGlyProLysSerSerTrpAlaProHisLysTrpSerThr 417
 |||||
 1412 ATTTCGAATAATACAGCAAGATGATTCGTGATTTAATCTATTCACCTANG 1461
 |||||
 417 rPheAlaAsnIleSerLysAspSerAspLeuIleLysSerThrTyG 434
 |||||
 1462 GGGAAAGACTGTGATCTCCCAAGTCATTCACGATCCCATGAGTTTGGCC 1511
 |||||
 434 LysLysAspSerAspLeuProSerAspPheSerIleHisGlnPheLeuAla 450
 |||||
 1512 ACGTGCCAAGATTAATCCGATATGTCATGCGAAGATAGTTTACTGATGTTT 1561
 |||||
 451 ThrLysGlnAspArgProLysValMetAlaAspSerLeuLeuAspValle 467
 |||||
 1562 AACAAAAGAGGGCATTCAGAGACCTTACAGAGATGGAAGATGTCATTCG 1611
 |||||
 467 uThrLysGlyLysHisSerArgThrLeuGlnGlnLeuMetGlnMetSerLeuP 484
 |||||
 1612 CTGAAGATGAAGGCCATPACTAGGACACTTGACAGAGAAAGAAATGGAG 1661
 |||||
 484 rGluAspGlnGlyHisThrArgThrLeuAspThrAlaLysGlnMetGln 500
 |||||
 1662 CAGATTACGAAGTAGAGCCACAGGGCGTTTGGACTCCAGTACTCAAGA 1711
 |||||
 501 ...lIethrGlnValGlnProProGlyArgLeuAspSerSerThrGlnAs 516
 |||||
 1712 CAGGCTCATACGCTGGAAGAGCATACAAATTTGGGTTCCACTGANG 1761
 |||||
 516 pArgLeuIleAlaLeuLysAlaValThrAsnPheGlyValProValGluVal 533
 |||||
 1762 TTTTGAAGCTGGAAGAACTGAATATTCAGAAAGAACTTGATGAGACC 1811
 |||||
 533 alphaAspSerGlnGlnLysAlaGlnIlePheGlnLysLysLeuAspGluThr 549
 |||||
 1812 ACCAGATTGCTCAGGAACTCAGAGAACCCAGAGATGAACCTTTGAGACAC 1861
 |||||
 550 ThrArgLeuLeuArgGlnLeuGlnGlnLysAlaGlnAsnGlnLysLeuSerTh 566
 |||||
 1862 CAGACCCCTGGGACATGATCTGTCTTGGGTCCTCATC.AGAGAA 1910
 |||||
 566 rArgProProAsnMetIleCysLeuLeuGlnProSerLysrArgGln 583
 |||||
 1911 TGCATCTTGTGTAACAGTACCAATTAATCTTAAGAATT.GCACAGCA 1959
 |||||
 583 eHisLeuAlaGlnGlnValThrAsnAsnLeuLysGlnLeuAlaGlnGln 599
 |||||
 1960 GTAATCCCAAGTGAATGTAAGTATGAGAGCTTGAAGAAAGCAATGGG 2009
 |||||
 600 ValThrProGlyAspIleValSerThrTrpGlyValArgLysAlaMetGln 616
 |||||
 2010 GATTTCATCTCTCCCGCTCATGGAAGAAACACTTTGTGATTTGACAG 2059
 |||||
 616 yLieserIleProSerProValMetGlnAsnAsnPheValAspLeuThrG 633
 |||||
 2060 AAGACACTGAAGAACCTAAAGAGAGGATGTGATGAGTGTGACCTGCT 2109
 |||||
 633 LysAspThrGlnGlnProLysLysThrAspValAlaGlnCysGlyProGly 649
 |||||
 2110 GGAAGT 2115
 |||||
 650 GlySer 651

seq_name: sp_human:Q9UH59

seq_documentation_block:

ID Q9UH59 PRELIMINARY; PRT: 652 AA.

AC Q9UH59; 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE BROMODOMAIN PROTEIN CELTIX1.


```

484 roGlaspGlUgLIhYsthrArgThrLeuAspThrAlaLysGluMetGlu 500
1662 CAGATTACAGAGTAGAGCCAGGCGGTTGGACTCCAGTACTCAGA 1711
501 GlnIleThrGluValGluProProGlyArgLeuAspSerSerThrGln 517
1712 CAGGCTCATAGCGCTGAAAGCAGTAACAAATTTGGCGTTCCAGTTGAAG 1761
517 pArgLeuIleAlaLeuLysAlaValThrAsnPheGlyValProValGlu 534
1762 TTTTGGAGCTGAGAGAGCTGAAATATTCCAGAGAACTTGTAGAGACC 1811
534 aPheAspSerGluGluValGluIlePheGlnLysLysLeuAspGluThr 550
1812 ACCAGATTGCTCAGGGAATCCAGAGAGCCCAAGATGAACTTTGAGCAC 1861
551 ThrArgLeuLeuArgGluLeuGlnGluAlaGlnAsnGluArgLeuSer 567
1862 CAGACCCCTGGGAACATGATGTGTCTCTGGGTCCTCATC.AGAGAA 1910
567 rArgProProAsnMetIleCysLeuLeuGlyProSerTyrArgGlu 584
1911 TGCATCTTGCAGACAGTACCAATTAATCTAAAGAAAT.GCAGAGCA 1959
584 eHisLeuAlaGluGlnValThrAsnAsnLeuGluLeuAlaGlnGln 600
1960 GTAATCCAGTATATGCTAAGACGATAGAGTTCGAAAGCAATGGG 2009
601 ValThrProGlyAspIleValSerThrTyrGlyValArgLysAlaMetGlu 617
2010 GATTTCCATTCCTTCCCGCTCATGGAAAAACAATTGTGGATTGGACAG 2059
617 yLleSerIleProSerProValMetGluAsnAsnPheValAspLeuThrG 634
2060 AAGACATGAGAACCTAAAAAGCAGAGTGTGATGAGTGCGACTGGT 2109
634 lAspThrGluGluProLysLysThrAspValAlaGluCysGlyProGly 650
2110 GGAGCT 2115
651 GlySer 652
seq_name: sp_rodent:088665
seq_documentation_block:
ID 088665 PRELIMINARY; PRT: 651 AA.
AC 088665;
DT 01-NOV-1998 (Tremblrel.08, Created)
DT 01-NOV-1998 (Tremblrel.08, Last sequence update)
DE 01-DEC-2001 (Tremblrel.19, Last annotation update)
DE BROMODOMAIN-CONTAINING PROTEIN BP75.
GN BRD7.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RP STRAIN-BALB/C; TISSUE-BRAIN;
RX MEDLINE=99456832; PubMed=10526152;
RA Cuppen E., van Ham M., Peeters B., Wieringa B., Hendriks W.:
RT bromodomain-containing protein."
RL FEMS Lett. 459:291-298(1999).
DR MGD: MGI:1349766; Brd7.
DR InterPro: IPR001487; Bromodomain.
DR Pfam: Pf00439; Bromodomain; 1.
DR SMART: SM00297; BROMO; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
SQ SEQUENCE 651 AA; 74000 MW; 5D34B4F14FD51350 CRC64;

```

```

alignment_scores:
Quality: 2966.50 Length: 652
Ratio: 4.701 Gaps: 1
Percent Similarity: 96.779 Percent Identity: 87.117
alignment_block:
US-09-687-230-1 x 088665 ..
Align seg 1/1 to: 088665 from: 1 to: 651
162 ATGGGCAAGAAAGCACAAGAGCAAGTCGGCAACACCTCTACAGAGA 211
1 MetGlyLysLysHisLysLysHisLysSerAspArgHisPheTyrGluGlu 17
212 GTATGTAGAGAAAGCCCTTGAAGCTGCTCTCCAACTAGAGAGGAAG 261
17 uTyrValGluLysProLeuLysLeuValLeuLysValGlySerGlu 34
262 TCACCGAACTCTCCACGGGACGCTGGGGACGACTCCAGCCTCTTCGA 311
34 aLThrGluLeuSerThrGlySerSerGlyHisAspSerSerLeuPheGlu 50
312 GACAAACGATCATGACAAACACAAGGACAGAAAGCGAAAGAAAGAGA 361
51 AsparGSerAspHisAspLysHisLysAspArgLysArgLysArgLys 67
362 GAAAGAGAGAAAGCAGATTCCAGGGGAAAGAAAGGAGAAACGAGAA 411
67 sLysGlyGluLysGlnAlaProGlyGluGluLysGlyLysArgLysArg 84
412 GAGTTAAGAGAGATTAAGAAAGAGAGATCGAGACCGGCTGAGAAATAG 461
84 rGValLysGluAspLysLysLysArgAspArgAspArgAlaGluAsnGlu 100
462 GCAAGAAAGATCTCCAGTGTACGCGCCCTGTGAGATTAGACTTCCCTC 511
101 ValAspArgAspLeuGlnCysHisValProIleArgLeuAspLeuPro 117
512 TGAGAAGCCTCTCACAGCTCTTATGACCAACAAAGAAAGTAGAAGA 561
117 oGluLysProLeuThrSerSerLeuAlaLysGlnGluValGluGlnT 134
562 CACCCCTTCAGAACCTTTGAATCAACTGATGAGCAATTGCAGAGAAA 611
134 hrProLeuGlnGluAlaLeuAsnGlnLeuMetArgGlnLeuAlaArgLys 150
612 GATCCAAAGCTTCTTTCATTTCTGAGACTGATTTATTTATGCTCCGG 661
151 AsproSerAlaPhePheSerPheProValThrAspPheIleAlaProGlu 167
662 CTACTCATGATCATTAACACCAATGATTTAGTACCATGAAGAAA 711
167 yTyrSerMetIleIleLysHisProMetAspPheSerThrMetLysGlu 184
712 AGATCAAGAACAACTATCATCGTCCATAGAGAACTTAAGATACATTC 761
184 ySileLysAsnAsnAspTyrGlnSerIleGluGluLeuLysAspAsnDhe 200
762 AAATAATGTACTAATGCCATGATTTACAATAACAGAGACCATTTA 811
201 LysLeuMetCysThrAsnAlaMetIleTyrAsnLysProGluThrIleTyr 217
812 TTATAAGCTGCAAGAGCTGTTCACCTCAGAAATGAATAATCTTACCC 861
217 rTyrLysAlaAlaLysLysLeuLeuHisSerClyMetLysIleLeuSerG 234
862 AGGAAGAAATTTAGAGCCTGAAGCAGAGCATRAGATTTAGCTACTTG 911
234 lngLysArgLysGlnSerLysGlnSerIleAspPheMetSerAspLeu 250
912 CAGAAACCTCGAAAGCAGAAAGATGGAACAGACCTCAGAGAGTGGGA 961
251 GlnLysThrArgLysGlnLysGlnLysArgThrAspAlaCysGlnSerClyGlu 267

```


53 PHEMETHALASPLEUGINLSTHARGLSGLNLSASPGLYTHRASPETH 69
 947 CTCACAGAGTGGGAGAGACGGAGCTGTCGACAGAGAGAGAGGAGACT 996
 69 ISEKINSERGLYGLNLSAPGLYGLYCYSRTPGLNARGGLNARGGLNLSAPS 86
 997 CTGGAGATGCCGACACACAGCTTCAAGAGTCCAGCAAAAGAAATATAA 1046
 86 ERLYASPLAIGLUNALHLSALAPHELYSERPROSERLGLUNLSNLS 102
 1047 AAGAAAGCAAAAGATATGCTTGAGATAGTTTAAACCAATATTTAGA 1096
 103 LYSLYSASPLYSASPMELEUGLUNSPLYSPHELYSSERASNLSNLEUGL 119
 1097 GAGAGAGCGAGAGACCTTGACCGCATGCTGAAGAAATCTGGAGGAAAGC 1146
 119 UNRGLGLNGLNGLNLEUNLSAPRGLLEVALYLSGLUSERGLYGLYSL 136
 1147 TGACACAGCGGCTTGTAAGACATGCGAATTTGAAAGAAAGAAACCA 1196
 136 EUTHRARGHARGLEUNVALNSERGLNCSGLUPHEGLNARGHARGLYSPRO 152
 1197 GATGGAACAGACAGCTTGGGACTTCTCCATCCTGTGGATCCCATTTGAG 1246
 153 ASPGLYTHRTHRTHLEUGLLEUNHLSPROVALASPROILEVALGL 169
 1247 AGAGCCAGCTACTGCTGCTGAGACTGGAGATGCAACTGAGAGACTTC 1296
 169 YGLUPROGLYTYRCYSPROVALARGLEUGLMEETHRTHRGLYARGLEUG 186
 1297 ACTCTGAGTGAATATCTTGACAGGGGTTCAAAAGANAAAAGAACANA 1346
 186 LNSERGLYALNSNTHRLEUGLNGLYPHELYSGLUNSPLYSRGASNLS 202
 1347 GTCACCCGAGTGTATATTTGATTTGAGCCCTACAGTTCTTATGACCC 1396
 203 VALTHRPROVALLEUNYTHREUNSNLYRGLYPROTYRSETTYRTHALR 219
 1397 GCATATGACTCCACATTTGCAATATACAGCAAGATGATTCGATTAA 1446
 219 OHLSYRASPSETHRPHALASNLSERLYSASPSPSERASPLEUL 236
 1447 TCATATCAACCTATGGGAAAGCTGATCTTCCAGTAGTTTACACATC 1496
 236 LEYTHRTHRTHRGLYGLNLSAPSERASPLEURROSERASPLESERIE 252
 1497 CATGAGTTTGGCCACGTGCCAAGATTATCCCTATGTCATGGCAGATAG 1546
 253 HISGLUPHEUNLATHRCYSGLNASPTYRPROTYRVALMETALASPE 269
 1547 TTTTACGTGATGTTTAAACAAAGAGGAGCTTCAGAGACCTTACAGACA 1596
 269 RLEUNLNASPVALLLEUNHRLYSGLYLHLSERATGTHRLEUGLGLN 286
 1597 TGGAGATGATATGCTGGAAGATGAAGGCCATCTAGACACTTGCACACA 1646
 286 ECTGLMETSERLEUPROGLNLSAPGLNGLYHLSRTHRTHRTHREUNSPETH 302
 1647 GGAAGAAAGATGAGACAGATTACAGAGTACAGCCACAGGCGCTTGA 1696
 303 ALALYSGLUNMETGLNGLNLETHRGLUNLGLUPROPROGLYARGLEUN 319
 1697 CTCACAGTACTACAGACAGCTCATAGCGCTGAAGACAGTAAACAAATTTG 1746
 319 PSETHRTHRTHRGLNLSAPRGLNLEALALEUNLSALVALTHRTHRASHPHEG 336
 1747 GCGTTCCAGTTGAAGTTTGAAGTCTGAAAGACCTGAAATATTTCCAGAA 1796
 336 LYVALPROVALGLUNLPHASPSERGLUNLALAGLUNLIEPHEGLNLS 352
 1797 AAACCTGATGAGACACAGATTGCTCAGGAACTCCAGAGACCCAGAA 1846
 353 LYSLEUNSPGLUTHRTHRARGLEUNLARGLNULEUNGLNGLNLS 369

1847 TGAACGTTTGAGACACAGACCCCTGGGAACATGATCTGCTTGGGTC 1896
 369 NGLNARGLEUNSERTHRARGPROPROPROASMETLIECYLLEUNGLY 386
 1897 CCTCATC.AGAGAAATGATCTTCTGTAACAGTGCACCAATATCTTAA 1945
 386 ROSETTYRARGLMEHLISLEUNLGLNGLNVALTHRTHRASNLEUNLS 402
 1946 GAATT.GCAGCAAGTATACCTCAGTGTATGATGTAAGCAGTATGAGT 1994
 403 GLUNLALAGLNGLNLVATHRPROGLYASPLILEALSERTHRTHRGLYVA 419
 1995 TCGAAAGCAATGGGATTTTCCATTCCTTCCCGCATGGAAGAAACACT 2044
 419 LARGLYSALMETGLYLSERILEPROSERPROVALMETCUNLSASN 436
 2045 TTGTGATTTACAGAGACACTGAAGACCTTAAAGACGAGTGTAT 2094
 436 HEVALASPLEUTHRGLNLSAPRTHRGLNGLUPROLYSLYTHRASPVALALA 452
 2095 GAGTCTGACCTGGTGAAGT 2115
 453 GLUCYSGLYPROGLYGLYSER 459
 seq_name: sp_human:Q96KA4
 seq_documentation_block:
 ID Q96KA4 PRELIMINARY; PRT; 351 AA.
 AC Q96KA4;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CDNA FLJ114402 FIS, CLONE HEMBA1003783, MODERATELY SIMILAR TO MUS
 DE MUSCULUS BROMODOMAIN-CONTAINING PROTEIN BP75 MRNA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP TISSUE=EMBRYO, AND MAINLY HEAD;
 RC Isoeqal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Nagatsuma M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
 RA Niromiya K., Iwayanagi T.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK027308; BAB55031.1; -
 SO SEQUENCE 351 AA; 39236 MW; 51E1DD2AB07DEDA0 CRC64;

alignment_scores:

Quality: 1716.50 Length: 352
 Ratio: 4.975 Gaps: 1
 Percent Similarity: 96.011 Percent Identity: 97.727

alignment_block:

US-09-687-230-1 x Q96KA4 ..

Align seg 1/1 to: Q96KA4 from: 1 to: 351

1062 ATGCTTGAAGATTAAGTTTAAAGCAATATTTAGAGAGAGACAGAGACA 1111
 1 MetLeuGLNAspLYSPheLYSSerASNLeuGLNArgGLNGLNGL 17
 1112 GCTTGACCCGATGCTGAAGAAATCTGAGAGAAAGCTGACAGCGGCTTG 1161
 17 nLeuAspArgLLeValLYSGlNserGLYLYSLeUTHrArgArgLeuV 34

```

1162 TGAACAGTCAGTGGCAATTTGAAGAGAAAACAGATGGACAAACGACG 1211
|||||
34 aIAsnSerGIncySgluPhegluAgluYlsPProAspEluThrlThr 50
1212 TTGGAGCTTCTCCATCCTGTGGATCCCATTTAGAGAGCCAGGCTAGTG 1261
|||||
51 LeuGluLeuLeuHISProValAsPProIleValIGluGluProGluYlTyCy 67
1262 CCTGTGAGACCTGGGAAATGACAACTGGAAAGACTTCACTGTGGAGTA 1311
|||||
67 sProValAgluLeuGluYlMetThrThrlYlAgluLeuInSerGluValAsnT 84
1312 CTTTTCAGGCGGTTTCAAGAGAGATAAAGAAACAAAGTCACTCCAGTTA 1361
|||||
84 hLeuGluInGluYlPheYlSgluAspYlsAgluAsnYlValThlProValLeu 100
1362 TATTGAATTAATGGCCCTACAGTTCTTATGCACCCGATTAATGACTCCAC 1411
|||||
101 TylLeuAsnTylGlyProTylSerSerTylAlaProHISlylAspSerTh 117
1412 ATTTCGAATATCAGCAAGAGATGATTCGATTTAATCTATTCAACCTATG 1461
|||||
117 rPheAlAsnIleSerYlsAspAspSerAspLeuIleTylSerThrlTyG 134
1462 GCGAAGACSTGTATCTCCAAAGTGAATTCAGATCCATGATTTTGGCC 1511
|||||
134 YgluAspSerAspLeuProSerAspPheSerIleHISgluPheLeuAla 150
1512 AGCTCCAAAGATTAATCCGTATGTCATGGCAGATTAATTTACTGATGTTT 1561
|||||
151 ThrCySgluAspTylProTylValMetAlaAspSerLeuLeuAspValLe 167
1562 AACAAAGAGAGGCGCATTCAGAGACCTTCAAGAGATGAGATGTCATTGC 1611
|||||
167 uThrYlSgluYlHISerAglThlGluGluMetGluMetSerLeuP 184
1612 CTGAAGACGAAGGCGCATTAAGACACTTGCACACAGAAAAGAAATGAG 1661
|||||
184 roglYAspGluGluYlHISThrAglThrlLeuAspThlAlaYlSgluMetLu 200
1662 CAGATTACAGAAATGAGCCACAGAGCGGTTTGACCTCCACTACTCAAGA 1711
|||||
201 ..IleThrlGluValIGluProGluYlAspSerSerThrlGluAs 216
1712 CAGGCTATAGCGCTGGAAGAGATTAACAATTTGGCTTCCAGTTGAG 1761
|||||
216 PArgLeuIleAlaLeuYlsAlaValThlAsnPhelYlAlProValIGluV 233
1762 TTTTGTACTCTGAAGAACCTGAATATTCAGAAAGAACTTGATGAGACG 1811
|||||
233 alPheAspSerGluGluAlaGluIlePheGluInYlsYlLeuAspGluThr 249
1812 ACCAGATGCTCAGAGAACTCCAGAGAGCCCAAGATGAACTTTGAGACG 1861
|||||
250 ThrArgLeuLeuArgGluLeuGluInGluAlaGlnAsnGluArgLeuSerTh 266
1862 CAGACCCCTGGGAACATGATCTGTCTGTGGTCCCTCAGC AGAGAAA 1910
|||||
266 rArgProProProAsnMetIleCySLeuLeuGluYlProSerTylArgGluM 283
1911 TGCATCTTGTGAACAACTGACCAATATCTTAAAGAAATTT GCACAGCAA 1959
|||||
283 eThISleuAlaGluGluInValThlAsnAsnLeuYlsGluLeuAlaGluIn 299
1960 GTAACCTCAGAGTATGCTGAAGACAGATGAGATTCGAAAAGCAATGGG 2009
|||||
300 ValThrProGluYlAspIleValISerThrlYlValAlaGluYlsAlaMetG 316
2010 GATTTCATCTCTTCCCTCCATGAGAAAACAACTTTGTGATTTGACAG 2059
|||||
316 YlISerIleProSerProValMetGluAsnAsnPhelValAspLeuThrg 333
2060 AAGACACTGAAGAAACCTTAAAGACGATGTTGATGAGTGTGACCTGTG 2109

```

```

seq_name: sp_invertebrate:09VLX2
seq_documentation_block:
ID 09VLX2 PRELIMINARY: PRT: 861 AA.
AC 09VLX2:
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG7154 PROTEIN.
GN CG7154.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasco P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Martel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Modarri C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
RA Switskas R., Tector C., Turner C., Turner E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003618; AAF52557.1;
DR Flybase: FBgn0031947; CG7154.
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; bromodomain.1.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO.1.
DR PROSITE: PSS0014; BROMODOMAIN_2; 1.
SQ SEQUENCE 861 AA; 95922 MW; E27D528E5F9B3A3F CRC64;

```

alignment_scores:

Quality: 738.50 Length: 824
 Ratio: 1.656 Gaps: 24
 Percent Similarity: 54.126 Percent Identity: 28.034

alignment_block:

US-09-687-230-1 x 09VLX2 ..

Align seg 1/1 to: 09VLX2 from: 1 to: 861

```

168 AAGAACACAAAGACAGACAGTGGAC...AACACCTCTACGAGAGTA 214
    |||||||.....:|||||
5  LysLysHisLysLysAsnLysSerGluArgGluLysTyrGluGluTyr 21
215 T.....GTAGAAAGCCCTTGAGCTGGCC 240
    |
21 rSerGlnHisGlnAspProAlaGlnLeuGlnArgGluLysLeuLeu 38
241 TCAAAGTAGAGGAGACAGAAAGTCCAGCACTCTCCAGGGCAGCTGGGG 290
    |||||||.....:|||||
38 euLysValIglYSerAsnAlaThrProGluTyrSerAlaAsnSerPromet 54
291 CACGAC.....TCCAGCCTCTT 307
    ||
55 ValAspGlyGlyProProThrAlaAlaGluAlaMetMetSerProValPr 71
308 CGAAGACAAAGACATCATGACAAACAGACAGAAAGCGAAGAAAGA 357
    |||||||.....:|||||
71 oGluGluLeuGlnAspHisGlnGlyHisArgGluArgHisLysLysSerI 88
358 GAAAGAAAGAGAGAAAGACAGATTCCAGGGAGAAAGAGGAGAAAGCG 407
    :|||||.....:|||||
88 yLysLysLysLysLys.....LysAspArgGluLysLysHis 100
408 AGAAGAGTTAAGAGAGTAAAGAACGACAGATCGAGACCGGTGGAGAA 457
    :|||||.....:|||||
101 LysHisHisLysGlu...LysArgHisArgSerArgAspArgHisArgAs 116
458 TCGAGCGAAG.....AAGATCTCCAGCTGC 483
    :.....:|||||
116 pAlaGlySerAspGluAspMetMetAlaGlyAlaAspAspAlaAlaCys 133
484 AC..... 485
133 erGlyPheAlaProSerSerValAlaProProAlaAlaAspProAspSer 149
485 ..... 485
150 SerGlnAspGlyPheSerPheMetAspAspGlnSerGlnProLeuPr 166
485 ..... 485
166 oGluAsnIleLeuPhePheAlaGlyIleThrThrAspAsnSerProSerA 183
486 .....GCCCTGTGAGATTAGC..... 503
    |||||.....:|||||
183 snCysProValThrLysProIleAlaProArgLysLeuAspAspIleLeu 199
503 ..... 503
200 MetGlySerSerProAsnSerSerSerLeuGlnSerSerSerLeuGlyLe 216
504 .....TTGCCT..... 509
    |||||
216 uIleGlySerSerProThrLysProLeuProAspLeuLeuIleProSerP 233
509 ..... 509
233 roSerThrProGlyGlyAlaAsnSerLeuAsnAlaLeuThrProLysAla 249
510 .....CCTGAGAGCCTCTCACAAAGCTTTAGCCAAACAGAA... 548
    |||||.....:|||||

```

```

250 LeuGluAlaProLysThrProSerSerSerSerGluSerGlyArgGluPr 266
549 .....GAGTAGAACAGACACCCCTTCAGAG 576
266 oArgSerCysValIleuLysLysGlnGlnLysSerProLeuAsnLysL 283
577 CTTTGAAATCAACTGATGAGACAAATTGCAGAGAAAAGATCCAGTGCCTTC 626
    |||||.....:|||||
283 euLeuGlnHisIleuLeuArgPheLeuGluLysArgAspProHisGlnPhe 299
627 TTTTCATTTCCCTGACATGATTTATTCCTGCTGCTACCATGATCAT 676
    |||||.....:|||||
300 PheAlaThrProValThrAspAspMetAlaProGlyTyrSerSerIleI 316
677 TAAACACCCCAATGATTTAGTACCATGAAGAAAGAAAGATCAAGCAATG 726
    |||||.....:|||||
316 eSerArgPrometAspPheSerThrMetArgGlnLysIleAspAspHisG 333
727 ACTATCAGTCCATAGAGAACTAAAGATTAACCTCAAACTAATGTGTACT 776
    |||||.....:|||||
333 LuTyrThrAlaLeuThrGluPheThrAspAspPheLysLeuMetCysGlu 349
777 AATGCCATGATTTACAAATPAAACAGACACATTTTATTAAGCTGCAAA 826
    |||||.....:|||||
350 AsnAlaIleLysTyrAsnHisValAspHisValItyAsnLysAlaAla 366
827 GAAGCTGTTCACATCAGCAATGAAATTTCTTAGCCAGAAAGAAAT...C 873
    |||||.....:|||||
366 sArgLeuLeuGlnValIglYMetLysHisLeuGlnProGluAsnLeuMet 383
874 AAGGCTGAAGCAGACATAGACTTCATGCGCTGACTTCGACAGAAACTGA 923
    :|||||.....:|||||
383 rGSerLeuLysProLeuSerGlyTyrMetArgGluLeu.....Thr 396
924 AAGCAGAAAGATGAAACAGACACCTCACAGAGTGGGAGAGAGAGCTGG 973
    :|||||.....:|||||
397 AlaArgGlnLeuGlyPheGluLeuSerSerAsnAsp..... 408
974 CTGGCAGACAGAGAGAGAGAGCTGTGGAGAT.....G 1005
    :|||||.....:|||||
409 .metSerArgGluAsnAsnAspSerAlaAspGluGlyAlaSerThrGlyA 425
1006 CGAAGCAGACCCCTTCAAAGATCCACAGCAAGAAATTAAGAAAGAAC 1055
    |||||.....:|||||
425 IagLgluLuproArgThrProAlaGlnLeuGluGluGluLupArgLysArg 441
1056 AAGATATGCTTGAAGATTAAGTTTAAAGC.....AA 1087
    |||||.....:|||||
442 ThrLeuArgLeuGlnAsnAlaProLysThrHisPheGluProTyrValAs 458
1088 TAATTTTAGAGAGAGAGCAG.....GAGCAGCTTGACCGCATCGAAGG 1131
    :|||||.....:|||||
458 pAspLeuThrGlyGluGluIleLeuAlaGlnValAlaGlnAsnAlaIaGln 475
1132 AATCTGGAGAGAAAGCTGACAGCGGCTGTGAACAGTACGTGCGAATT 1181
    :|||||.....:|||||
475 ImlaLysGlnArgValAsnAlaLysLysAsnAlaHisLysMetGlyPhe 491
1182 GAAAGAGAGAAACACAGATGAACAACAGAGCTTGGCATTCATCTCTCT 1231
    |||||.....:|||||
492 LeuArgGlnMetLysAspGlyThrThrLeuAsnLeuValIleLysGln 508
1232 GATGCCCATTTAGAGAGACCCAGCTACCTGCTGGAGACATGGGAATGA 1281
    :|||||.....:|||||
508 uGlu.....AsnGluGlyProGluIuArgValAlaThrIleGlyAspL 522
1282 CAACCTGAGAGACTTCACTGTGAGTGAATATCTTGGAGGGGTTCAAGAG 1331
    :|||||.....:|||||
522 euValIglYLysLeuGlnLysGlySerAlaGlnLeuGlnThrArgGlnVal 538
1332 GATTAAGAAAGACAAAGTCACTCCAGTGTATTAATTAATTAATGAGCCCTTA 1381
    |||||.....:|||||
539 AspLysArgAsnAlaValArgThrValLysSerSerLeuAsnTyrGlyAlaPh 555

```

```

1382 CAGTCTTATGACACCGCATATGACTGACATTTGCAAAATATCAGCAGG 1431
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
555 ealaSerPhalaProthrPhaeSperSerrgPheSerThleuSerIaG 572
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
1432 ATGATTCGTATTAATCTATTATTCATGAGGAGAGACTGATCTTCCA 1481
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
572 IuGluThrGlnLeuValLeuArgThrTyroIyAspAlaSerIaGlu 588
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
1482 AGTGAATTCACATCCATGAGTTTGGCCAGCGCCAGATTAATCCGTA 1531
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
589 TyraIaGluSerIleuGlnPheThrLysAsp...SerSsnTyroIyTh 604
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
1532 TGTCATGCGCAGATAGTTTACTGATGTTTAAACAAAGAGGCGCATCCA 1581
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
604 rTrhIleAlaSnclLyuLeuAspIleuThrAsnGlyLuhHisSerI 621
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
1582 GGACCCCTA..... 1592
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
621 ySerIleuAspGluLeuTyraSnmMetGlnLeuHisSerTyroIyGluArg 637
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
1593 GAGATGAGATGTCATTCCTGGAAGATGAAGCCCATCTAGAGCACTTGA 1642
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
638 GluIleGluLysCysPheGluGlnGlu.....G1 648
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
1643 CACAGGAAAAAATGAGCAGATTAACAGAAATGAGCCACCGGCGTT 1692
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
648 urhSerSerGlnGlnIuThrAlaGlnIleGluGlnIuTyroIuL 665
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
1693 TGGACTCCAGTACACAGACAGGCTCATAGCGGTGAAGAGTAAACAA 1742
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
1743 TTTGGCGTTCAGTTGAAGTTTGTGACTCTGAAGAGCTGAATA.... 1787
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
682 leuGlyIleAspValSerPheLeuAspGlyMetGluIaGluMetLys 698
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
1788 .....TTCAGAAAGAACTTGATGAGACCCACCGAT 1818
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
698 rTyroIyLeuAsnArgArgMetHisGlnHisLeuSerGlnAsnLeuThrL 715
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
1819 TGGTCAGGAAATCCAGAAAGCCAGAAATGAGCTTGACACACGAGCC 1868
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
715 euIleGluLysLeuArgValAlaGlnHisAspArgLeuSer...GlnPro 730
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
1869 CTTGGGCAATGATCTGTCTGTGGGTCCCTCATC...AGAGAAATGCA 1914
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
731 leuProAsnHisLeuGlyLeuValGlnProAlaGlyGlnIuIleG1 747
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
1915 TCTTGTGCAACAGTGCACAAATATCTTAAGA.ATTGCACAGCAAGTAA 1963
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
747 nThraIaGlnGlnLeuThrGlnGlnIleSerAspLeuAlaLysLysLeuP 764
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
1964 CTCACAGTATATCTGAACAGCAGATGAGTGCAGAAAGCAATGGGAT 2013
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
764 roProSerAlaIleAlaAspProTyraIaLeuArgLysAlaMetGlyMet 780
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
2014 TCC.....ATTCCCTCCCGCCGTCATGAGAAACAACATTTGTGATTT 2054
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
781 SerTyraIaGlyLeuProProArgProValSerProArgValGlnLe 797
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
2055 GACAGAAAGACATGAAAGACCT 2076
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
797 uProGluLeuLeuGlnGlnPro 804
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
seq_name: sp_human:Q9H8M2
seq_documentation_block:
ID Q9H8M2 PRELIMINARY; PRT; 501 AA.
AC Q9H8M2;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)

```

```

DE CDNA FLJ13441 FIS, CLONE PLACE1002775, WEAKLY SIMILAR TO
DE PERGRIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y., Oshima A.;
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL: AK023503; BAB14591.1;
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; bromodomain.1.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
SQ SEQUENCE 501 AA; 55664 MW; 5C68F53097BA073C CRC64;

alignment_scores:
      Quality: 684.00      Length: 473
      Ratio: 2.178      Gaps: 11
Percent Similarity: 66.385      Percent Identity: 34.672

alignment_block:
us-09-687-230-1 x Q9H8M2 ..

Align seg 1/1 to: Q9H8M2 from: 1 to: 501

549 GAAGTGAACAGACACCCCTTCAAGAGCTTGATGAATCAATGATGAGACA 598
      ||| :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
19 GluAsnGluSerThrProIleGlnGlnLeuLeuGlnIuHisPheLeuArg1 35
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
599 ATTGCAGAGAAAGATCCAAAGTCTTTTCATTTCCGTGACATGAT 648
      ||| :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
35 nLeuGlnArgLysAspProHisGlyPhePheAlaPheProValThrAspA 52
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
649 TTTATGCTCGCTGCTACTCCATGATCATTTAAACACCAATGATTTTACT 698
      ||| :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
52 lAlaIleAlaProGlyTySerMetIleIleLysHisProMetAspPheGly 68
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
699 ACCATGAAAGAAAGATCAAGAACATGATCACTACGCTAGAGAGACT 748
      ||| :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
69 ThrMetLysAspLysIleValAlaAsnGluTyroLysSerValThrGluPn 85
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
749 AAAGATTAACCTCAACTAATGTGTAATGTCATGATGATTAACAAATAC 798
      ||| :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
85 elyAlaAspPheLysLeuMetCysAspAsnAlaMetThrTyraSnrArgP 102
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
799 CAGAGACATTTATTAAGCTGCAAGAGAGCTGTTCACCTAGAGAAATG 848
      ||| :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
102 roAspThrValTyroTyroLysLeuAlaLysLysIleLeuHisAlaGlyPhe 118
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
849 AAATTCCTTGGCCAGAAAGATTCAGAGCTGAGAGCATGAGTACT 898
      ||| :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
119 LysMetCysSerLysGluArgLeuLeuAlaLeuLysArgSerMetSerPn 135
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
899 CATGGCTGACTTCAGAAACATCGAAAGCAGAAAGATGAAACAGACACT 948
      ||| :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
135 emetGlnAspMet.....AspPheS 142
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
949 CACAGAGTGGGAGAGAGAGGCTGCTGCGCAGAGAGAGAGAGAGACTCT 998
      ||| :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
142 erGlnGlnAlaAlaLeuLeuGly.....AsnGluAspThr 153
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::
999 GGAGATGCCAGACACAGCGCTTCAAGAGTCCC.....AG 1033
      :::::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::

```

```

154 AlaValGluGluProValProValGluValGluThrAl 170
1034 CAAGAAATAAATAAGAAAGCAAGATATGCTGAGATAAGTTAA. 1082
      |||:|||||:|||||:|||||:|||||:|||||:
170 alyslsSerLysLysProSerArgGluValIleSerCysMetPheGluP 187
1083 .....AGCAATAAATTTAGACAGACAG 1103
      |||:|||||:|||||:|||||:|||||:
187 rogluGlyAsnAlaCysSerLeuThrAspSerThrAlaGluIleHisVal 203
1104 CAGGAGCAGCTGACCGCATGCTGAGAAATCGAGAGCAACGACGACG 1153
      |||:|||||:|||||:|||||:|||||:|||||:
204 LeuAlaLeuValGluHisAlaAlaAspGluAlaArgAspArgIleAsnArg 220
1154 GCGGCTGTGACAGCTAGTGGCAATTTGAAAGAAAGAAACCAAGTGGAA 1203
      |||:|||||:|||||:|||||:|||||:|||||:
220 gphleuProGlyGlyLysMetGlyTyrLeuLysArgAsnGlyAspGlyLys 237
1204 CACGACGCTGGGACTTCTCCATCTCTGTCGATCCATTTGAGAGACCA 1253
      |||:|||||:|||||:|||||:|||||:|||||:
237 erLeuLeuTyrSerValValAsnThrAlaGluProAspAlaAspGluGlu 253
1254 GGCTACTGCTGCTGAGACTGGCACTGACAACTGGAGACTTCAGCTGG 1303
      |||:|||||:|||||:|||||:|||||:|||||:
254 GluThrHisProValAspLeuSerSerLeuSerSerLysLeuLeuProGly 270
1304 AGTGAATCTTTCAGAGGGTTCAAAGAGATAAAGAAACAAGTCACTC 1353
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
270 yPheThrThrLeu...GlyPheLysAspGluThrArgAsnLysValThr. 285
1354 CAGTGTATTTGATTTAGTGGCCCTACAGTCTTATGACCCGATTTAT 1403
      |||:|||||:|||||:|||||:|||||:|||||:
286 .....PheLeuSer...SerAlaThrThrAlaLeuSerMetGlnSer 298
1404 GACTCCACATTTGCAATATATGCAAGAGATGATTCGATTTATATATTC 1453
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
299 AsnSerValPheGlyAspLeuLysSerAspGluMetGluLeuLeuTyrSe 315
1454 AACCTATGGGGAAGACTGATCTTCACAGATGATTCAGCATCATGAGT 1503
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
315 rAlaTyrGlyAspGluThrGlyValGlnCysAlaLeuSerLeuGlnGluP 332
1504 TTTCGGCCACGTCGCAAGATTCGCTATGTCATGAGCATGCTTACG 1553
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
332 heValLysAspAlaGlySerTyrSerLysValValAspAspLeuLeu 348
1554 GATGTTTAAACAAAGAGGCGATTCAGACCTCAAGAGATGAG.. 1601
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
349 AspGlnIleThrClyGlyAspHisSerArgThrLeuPheGlnLeuLysGly 365
1602 .....ATGTCATTGCTGAGAGATGAGAGCCCTACTAGCA 1635
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
365 narGArgAsnValProMetLysProProAspGluAlaLysValGlyAspT 382
1636 CACTT...GACACAGAAAGAAATGAGACAGATTACAGAAATGAGGCA 1682
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
382 hLeuGlyAspSerSerSerSerValLeuGluPheMetSerMetLys... 397
1683 CCAGGGCGTTGGACTCCAGTACTCAAGACAGCTCATAGCGCTGAAAC 1732
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
398 .....SerTyrProAspValSerValAspIleSerMet 408
1733 AGTAACAATTTTGGCGTTCAGTTGAAAGTTTTCATCTGAGAAAGCTG 1782
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
408 LeuSerSerLeuGlyLysValLysGluLeuAspProAspAspSerH 425
1783 AAATATTCAGAGAAACTGATGAGCACACAGATGCTCAGGGAAGCTC 1832
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
425 IsLeu.....AsnLeuAspGluThrThrLysLeuLeuGlnAspLeu 438
1833 CAGGAAGCCACAGATGACGTTTGACACACAGACCCCTGGAGACATGAT 1882
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
439 HisGluAlaGlnAlaGlnArgGlyLysArgProSerSerAsnLeuSe 455

```

```

1883 CTGTCTCTTGGGTCCTCA 1901
      |||:|||||:
455 rSerLeuSerAsnAlaSer 461

```

```

seq_name: sp.rodent.09CT78

```

```

seq_documentation_block:

```

```

ID 09CT78      PRELIMINARY;      PRT;      174 AA.
AC 09CT78;
DT 01-JUN-2001 (TREMBLERel. 17, Created)
DT 01-JUN-2001 (TREMBLERel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLERel. 19, Last annotation update)
DE 18 DAYS EMBRYO CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
DE CLONE:1190001E05, FULL INSERT SEQUENCE (FRAGMENT).
GN BRD7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
OX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Atakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kodate K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fletschmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderrelli R., Barsch G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitlaker C., Wilting L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohatsu S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK004429; BAB2329.1; -.
DR MGD; MGI:1349766; Brd7.
FT NON_TER 1
FT 1
SO SEQUENCE 174 AA; 19027 MW; CF47FE3B9D719D03 CRC64;

```

```

alignment_scores:

```

```

Quality: 635.50      Length: 175
Ratio: 3.972      Gaps: 1
Percent Similarity: 91.429      Percent Identity: 75.429

```

```

alignment_block:

```

```

US-09-687-230-1 x 09CT78 ..

```

```

Align seg 1/1 to: 09CT78 from: 1 to: 174

```

```

1593 GAGATGAGAGTCTCATTCGCGAAGATGAGAGCCCTACTAGACACTGGA 1642
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1 AspLeuAspMetSerSerProGlnAspGluGlnThrArgAlaLeuAs 17
1643 CACAGAAAGAAATGAGACAGATTACAGAAATGAGACAGGCGCTT 1692
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
17 pphrAlaLysGluAlaGlu...IleThrGlnIleGluThrProThrGlyArgL 33
1693 TGGACTCCAGTACTCAAGACAGCTCATAGCGCTGAAAGCAGTAAACAAT 1742
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
33 euGlnSerSerSerGlnAspArgLeuThrAlaLeuGlnAlaValAlaThrThr 49
1743 TTTCGCGTTCACGTTCAAGTTTTCGCTGAGAAAGCTGGAATATTCGA 1792

```


50 PhcglyAlaPProAlaGluValIheAspSerGluGluAlaGluValPheG1 66
1793 GAAGAAACTGTATGAGACACCCAGATGTGTCAGGAACTCCAGNAAGCC 18412
66 nargylsleuAaSpIuThrThrArgleuenuArgIuenuGluAlaG 83
1843 AGAATCAACGTTTGAGCACCAACCCCTGGGAAcATGATGTCTCTTG 18922
83 lAsnGluArgleuSerThrArgProProPAsmetIleCysleuenu 99
1893 GGTCCTCATC AGAGAAATGCATCTTCTGTAAACAAGTGACCATAACT 19411
100 GlyProSerTyrArgIuMetIytleuAlaGluGluValIThrAsnAsnIle 116
1942 TAAAGATTGCA CAGCAAGTAACTCAGAGTATGTCTGAACAGTATG 19900
116 uylsGluIeuThrGluGluAlaValThrProGlyAspValValSerIleHtSg 133
1991 GAGCTTCGAAACCAATGGGGATTTCATCTCTCCCGCTCATGGAAAC 20404
133 lValAlaArgysAlaMetGlyIleSerValIProserProIleValGlyAsn 149
2041 AACTTTTGCGATTTTACACGAAGACACCTGGAACACCTAAAGAGCGATGT 20900
150 SerPheValAspIeuThrGlyCylucysGluGluProIysGluThrSerTh 166
2091 TGATGAGTGTGCACCTGGTGGAAAT 2115
166 rAlaGluCylucysGlyProAspAlaSer 174

seq_name: sp_invertebrate:Q17581

seq_documentation_block:

LD	Q17581	PRELIMINARY;	PRT;	636 AA.
NC	Q17581			

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last seq

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE COLH6.7 PROTEIN.
CN COLH6.7

GN COLH6.7.

05 *Caenorhabditis elegans*.
06 *Eubacterium* *metastaticum* at 100

Eukaryota; Metazoa; Nematoda; Chromadorea

NCBI TaxID=6239

NCBTL_IAXLD=0239;
[1]
ON
BN

RP SEQUENCE FROM N A

SEQUENCE FROM N.A.
Berks M.

Submitted (APR-1996) to the EMRI/GenBank/DBIT database

AD [2]
RN

SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA	none;
RA	none;

RT "Genome sequence of the nematode *C.elegans*: A platform for

Investigating bio

Science 282:2012-2018 (19

EMBL; 271258; CAA95779.1; -.

DR InterPro; IPR001487; Bromod

DR Pfam; PF00439; bro

DR SMART; SM00297; BR

DR PROSTIE; PS50014;
SO SEQUENCE 635 AA.

SEQUENCE 636 AA; 71339 MW; 02A5B089B8DCD663 CRC64;

alignment_scores:

Quality: 517.50

Ratio: 1.387

Percent Similarity: 52.461 Percent Identity: 25.598

a) Harmoniz. b) Conf.

alignment_block: MS-09-697-330-1 " 017501

US-09-687-230-1

Aljan sec 1/1 to: 017591 from: 1 to: 236

Align seq 1/1

```

879 CTGAAGCAGATAGACTTC..... 899
900 .....ATGCTGACTTCAGAAAATCGA.....AGCAGA 930
272 ylleatgproleuhalaprovalprolysgluarhythmetasnllyargl 289
931 AAGATGGAACAGACACCTCAGAGAGTGGAGAGAGAGCTGCTGGCAG 980
289 yalalavalalysaspglymetthrSerluasp.....Cysleugln 303
981 .....AGAGAGAGAGAGACTCTGAGATCCGGAAGC 1012
304 ValAlaSProlysValArgGluArgLeuSerAlaLysLeuProGluAl 320
1013 ACACGCCCTTCAGAGTCCAGCAAAAGAAAATAAAAGAACCAAGATA 1062
320 a.....AsnAsnProLysAsnLysYm 328
1063 TCGTTGAAGATAGTTTAAAGCAATATTAGAGAGAGAGAGAGCAG 1112
328 etelylsleugllypneleuser..... 335
1113 CTTGACCCGATCCTGAAGAAATCTGGAGAAACCTGACAGCGGCTGT 1162
336 .....GluLysAspGlyThrValValLeuAsnValVa 346
1163 GAACACTGACTGCGAA.....TTTGAAGAAAGAAAACCAAGATGGA 1203
346 lalaglyaspSerGluaspGlyLysLeuGluAsnAlaProArgArg 363
1204 CAACGAGTGGGACTTCTTCATCTGTGGATCCATTGTAGAGAGAGCA 1253
363 alhrrtleaglyaspIleValGlyProLeuGluLuu.....GlyThrPro 377
1254 GGCTACTGCTGTGAGACTGGGAATGACAACTGGAAAGCTTACGCTGG 1303
378 Gly.....MetIleGlnMetAlaAspHisArgLeuPheSer... 389
1304 AGTGAATACTTTCAGAGGGTTCAAAAGAGTAAAGAACCAAGTCACTC 1353
390 .....GlnAlaLp 392
1354 CAGTGTATATTGAATTATGGCCCTACAGTTCTTATGACCCGATTAT 1403
392 rovalAsnTyrlLeuAsnTyrlGlyProTyrSerSerPheAlaPrometLyr 408
1404 GACTCCCACTTTCGAATATACAGCAAGAGATTCGATTATTCATTTC 1453
409 AsperthrThrPalatInrMetThrLysGluAspThrAspLeuPheLeuAr 425
1454 AACCTATGGGGAAGACTGTGATCTTCACAGTATTCAGCATCATGAGT 1503
425 gthrtTyrlGlyAspLysSerAsnAlaSerValMetSerMetArgArgp 442
1504 TTTTGGCCAGCTGCCAAGATTATCCGATATGTCATGAGCAAGTATTC 1553
442 hevalGlyAspCysProGluPheSerGluIleIleGly...SerLeuLeu 457
1554 GATGTTTTAAACAAAAGAGGAGCTTCAGAGACCTTACAGAGATGAGAT 1603
458 AsphthrLeuThrAspGlyGluHisSerLysThrMetLysGluLeuGlu.. 473
1604 GTCATTTGCTGAGATGAAAGGCATCTAGACACTTGACACAGGAAAG 1653
474 .....AsnAlaGlyLysG 478
1654 AAATGGAGCAATATACAGAACTAGAGCCACAGGCGTTGGACTCCAGT 1703
478 luvallValysGlu.....GluValAspAsnAspGlyTyrLysAsnGluThr 492
1704 ACTCAAGACAGGCTCATAGCGCTGAAGACACTAAACAATTTGGCGTTCC 1753

```

```

493 ValLeuSerLeuIleAspAspValSerSerIleSerAsnLeuGlyIleGl 509
1754 AGTGAAGCTTTTACCTCTGAAAGAGCTGAATATTC..... 1790
509 utrhGlyPheLeuAsnAspIleArgGlnGlnValLeuValProAlaValG 526
1790 ..... 1790
526 luserAsnIleGluAsnAsnIleProGluPheMetAsnGluValAsnHis 542
1791 .....CAGAAGAACTGATGAGACCCAGATTTGCTCAGGAGACT 1831
543 MetAsnValGlnGlnGlnLeuAsnHisSerGlyGlnLysValLysAspLe 559
1832 CCAGAGACCCCAAGATGAACGTTTGAAGCAGCAGACCCCTGGGAACATGA 1881
559 uAlaHisIleGlnGlnHisArgLeuValGlnGlnProProProMetIleM 576
1882 TCTGCTCTTGGGCTCCCTCATCAGA..... 1906
576 etSerValGlnGlyValGlyGlnIleGlnGlnLysLeuAlaGluAsnLeu 592
1907 GAATGCAATCTTGTGAGACAGTACCAATAT 1939
593 GlnGlnHisLeuAlaHisGlnMetThrThrHis 603

```

```

seq_name: sp_human:Q9H7R9
seq_documentation_block:
ID Q9H7R9 PRELIMINARY; PRT; 233 AA.
AC Q9H7R9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CDNA FLJ14330 FIS, CLONE PLACE400261, WEAKLY SIMILAR TO
DE PERGRIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.,
RT "NEDO human cDNA sequencing project."
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK024392; BAB14907.1; -.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; Bromodomain; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR PROSITE; PS50014; BROMODOMAIN-2; 1.
DR PROSITE; PS50014; BROMO; 1.
SQ SEQUENCE 233 AA; 26233 MW; FBF96245BD88F4EB CRC64;

```

```

alignment_scores:
  Quality: 391.00 Length: 106
  Ratio: 4.116 Gaps: 0
  Percent Similarity: 89.623 Percent Identity: 65.094

```

```

alignment_block:
US-09-687-230-1 x Q9H7R9 ..
Align seg 1/1 to: Q9H7R9 from: 1 to: 233

```

```

549 GAAGTGAACAGACACCCCTCAAGAGCTTGAATCACTAGATGAGACA 598
19 GluAsnGlnSerThrProIleGlnGlnLeuLeuAspHisPheLeuArgGl 35

```

```

599 ATGCAGAAAGAAATCCAGTCTTCTTTCTTCTTCTGACTGATT 648
|||||
35 nleuglnarlgysaprophisglpnehealaphneprovalthrpsa 52
|||||
649 TTATTCCTGCTGCTACTCATGATCATTTAAACCAACCATTTAGT 658
|||||
52 lailealaprogllyrsermetlelelyshsiprokelsaphegly 68
|||||
699 ACCATGAAGAAAGATCAAGAACATGACTATCATCTCATAGAACAT 748
|||||
69 ThmetelysapslyilevalalaansglutrylyserValthrgh 85
|||||
749 AAAGATAACTTCAACTAATGTGACTAATGCCATGATTACATTAAC 798
|||||
85 elysalalaspheleyleuemetccysapsnalamekthrlyrasnar 102
|||||
799 CAGAGACCATTTTATTAAGCTCAAGAACGCTGTGCATCAGGAAG 848
|||||
102 rospthryVallyrlyrlyslleualalyelyleuleuhsialagly 118
|||||
849 AAATCTTAGCCAGCA 866
|||||
119 lysmetleuserlysgln 124
|||||

```

seq_name: sp_human:Q9H505

```

seq_documentation_block:
ID Q9H505 PRELIMINARY; PRT; 362 AA.
AC Q9H505.
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CDNA: FLJ23177 FIS, CLONE LNC10649.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Oho T.,
RA Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohmori Y., Ota T., Suzuki Y.,
RA Oobayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA Isogai T., Sugano S.;
RT "NDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK026830; BABI5565.1;
SQ SEQUENCE 362 AA; 39411 MW; 3B8240D7C70E44A5 CRC64;

```

alignment_scores:

Quality:	274.00	Length:	307
Ratio:	1.420	Gaps:	8
Percent Similarity:	62.866	Percent Identity:	25.733

alignment_block:

US-09-687-230-1 x Q9H505 ..

Align seg 1/1 to: Q9H505 from: 1 to: 362

```

1032 AGCAAAAGAAATTAAGAAAGCAAGATGCTTGAAGATAAGTTAA 1081
|||||
31 Alalylslyserlyslsproserarlgluvalilesercysmetphei 47
|||||
1082 A.....AGCAATTAATTTGAGAGAG 1101
|||||
47 uprogluglyasnalacysseurthrpserserthrAlaglugluhiv 64
|||||
1102 AGCAGAGACGCTTGACCGCATGCTGAAGAAATCTGAGGAGAAAGCTGACC 1151
|||||
64 alleualaleuValgluhtlealalalaspriunlalargaparvlgilean 80
|||||

```

```

1152 AGCGGCTTGTCAGACGATCGATGCCAATTTGAAAGAGAAACAGATGG 1201
|||||
81 Argpneleuproglyglyllysmetgllyrlyleuysarlgasnglyasp 97
|||||
1202 AACAAAGACGCTTGAGATTCATCTGATGATGCCATTTAGAGAGGC 1251
|||||
97 yserleuenuyrserValvalasnthAlagluProspAlasplug 114
|||||
1252 CAGGCTACTGCCGTGGTGAAGTGGAAATGACATGGAAGACTTCAGTCT 1301
|||||
114 luguThrhisProvalaspserSerleuserSerlyleuenuPro 130
|||||
1302 GGAGTGAATACCTTTGACAGGCTTCAAGAGATTAAGAAACAAAGTCAC 1351
|||||
131 Glypnehrthrleu...GlypnehsapsgluargasnllyValth 146
|||||
1352 TCACATGTTATTTGATTTGAGGCTTACAGTCTTATTCACCGCAT 1401
|||||
146 r.....pneuser...SerAlaThrThrAlaleusermetcIna 159
|||||
1402 ATGACTCCACATTTGCCAATATTCAGCAAGATGATTCATTTAATCTAT 1451
|||||
159 snasnservalpneglyaspneulyssearpsglumetgluleuenuyr 175
|||||
1452 TCAACCTTAGGGGAGACCTGATCTTCCAGATGATTCAGACATCCATGA 1501
|||||
176 SerAlaThrlyrlyaspgluthrlyValglncysAlaleuserleuIngl 192
|||||
1502 GTTTTGGCCAGCTGCCAAGATTTCCGTATGTCATGAGATAGTTAC 1551
|||||
192 upheVallyslapslaaglyserlyserlyslsValvalaspneul 209
|||||
1552 TGCATGTTTAAACAAAGAGGAGGATCCAGACCCCTACAGAGATGGAG 1601
|||||
209 euaspglInlethrlyglylshsiserarthrleupneuglnleuys 225
|||||
1602 .....ATGTCATTCCTGAAGATGAGAGCCATCACTAG 1633
|||||
226 GluArgArgasnlValPrometllyspProProaspgluAlalysValalys 242
|||||
1634 GACACT...GACACAGAAAGAAATGGACGAGATTCAGAGATGAGC 1680
|||||
242 pthrleuglyaspsersearserlyValleugluPneuserMetlyls. 258
|||||
1681 CACACAGGCGCTTGACACTCCAGATTCACAGACAGCGCTCATAGCCGTA 1730
|||||
259 .....SetTyrProhspsValserValaspIleSer 268
|||||
1731 GCAGTACAAATTTGGCGTCCAGTTGAAGTTTGGACTGGAAGAGC 1780
|||||
269 MetleuserSerleuenglyslsVallyslsgluLeuaspProaspse 285
|||||
1781 TGAATATTCAGAGAAAGAACTGATGAGACCAACGAGATTCACAGGAAC 1830
|||||
285 rhtsleu.....AsnleuaspgluthrlyslsleuInglasp 299
|||||
1831 TCCAGAGACCCAGATGAACGTTTGAGACCAAGACCCCTGGAGAACATG 1880
|||||
299 euhlsleuInglInalIngluArgglylserArgProserSerleuenu 315
|||||
1881 ATCTGTCTTGGTCCCTCA 1901
|||||
316 SerSerleuserAsnAlaser 322
|||||

```

seq_name: sp_human:Q9Y403

```

seq_documentation_block:
ID Q9Y403 PRELIMINARY; PRT; 715 AA.
AC Q9Y403.
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)

```


"Sequences cloned from human fetal thymus cDNA library."
 RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF167351; AAD4658.1; "
 SQ SEQUENCE 56 AA; 6396 MW; 431E06FC3D9D7430 CRC64;

alignment_scores:

Quality: 252.00 Length: 55
 Ratio: 4.846 Gaps: 0
 Percent Similarity: 94.545 Percent Identity: 85.455

alignment_block:

US-09-687-230-1 x Q9SN19 ..

Align seg 1/1 to: Q9UN01 from: 1 to: 56

1266 GTGAGCTGGTGAACACTGAGAGCTTACAGTTCAGTGAATACACTT 1315
 1 MetArgLeuGlyMetThrThrGlyArgLeuGlnSerGlyValAsnIleLe 17
 1316 GCAGCGCTTCAAGAGCATTAAGACAAAGTCACTCCAGTCTTAATAT 1365
 17 uGlnGlyPheLysGlnAspLysArgAsnLysValThrProGlyLeuTyrL 34
 1366 TGAATTAATGGCCCTACAGTTCTTATGACCGCATTTAGACTCCACAT 1415
 34 euAsnTyrGlyProTyrPserSerTyrValProHisTyrAspSerThrPhe 50
 1416 GCAATATTCAGCAAG 1430
 51 MetAsnIleLysGln 55

seq_name: sp_plant:Q9SN19

seq_documentation_block:

ID Q9SN19 PRELIMINARY; PRT: 952 AA.
 AC Q9SN19;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, last annotation update)
 DE SIMILAR TO DJ522J7.2.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OC NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 3, PAC
 RT clone:PO043E01."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF000615; BAA85417.1; "
 DR InterPro: IP001487; Bromodomain.
 DR Pfam: PF00439; bromodomain; 1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR PROSITE; PS0014; BROMODOMAIN_2; 1.
 SO SEQUENCE 952 AA; 105312 MW; 91D8E06A5A1769D CRC64;

alignment_scores:

Quality: 244.50 Length: 802
 Ratio: 0.745 Gaps: 40
 Percent Similarity: 40.898 Percent Identity: 22.195

alignment_block:

US-09-687-230-1 x Q9SN19 ..

Align seg 1/1 to: Q9SN19 from: 1 to: 952

42 CCTGGGGGGGGGTCGCGGGGGCCGCGCTCCGCTGGCTG 91

||||| ||| ::::: ||| |||||
 20 ProArgHisAlaValAspArgSerHisThrSerProProPro..... 33
 92 GCCCGACCGCGAGACGGGGCGGACGCGGCTGGGCTCGGCGC..... 132
 34ProProAsnAlaProAlaArgGlyArgProIle.ArgAsnArgLeu 47
 133 ...GGGGGGCGGACCGGGGCCCGGTCGACATCGGCAAGACAGCAAG 179
 48 HisGlyAlaLeuGlnProAsnProAlaSerMetProAlaProArgGly 64
 180 AAGCAATGTCGACAAACACCTCTACAGAGATGTATAGAGAACCCCT 229
 64 SAsnThrLysArgThrGlnSerHisThrHis.....Pro. 75
 230 GAACCTGCTCCTCAAGTAGAGAGGAGGAGGAGTCCAGCTTCAC.. 277
 76 AspProPheProAlaLysAlaThrProArgProHisHisHisHisH 92
 277 277
 92 sLeuLeuLeuProProProProArgProSerMetAlaLysThrArgL 109
 277 277
 109 ySaAlaAlaAlaProProProProProProProAlaGluThr 125
 278GGGACGCTCGGGGACGACTCCAGCTTCGAG 312
 126 ProAlaArgArgLysGlyLysLysLysGlyArgProSerLeuLeuAsp 142
 313 ACAAAGCATCATGACAAACAAAGACAGCAAGCCGNA.....AAG 356
 142 uGlnArgArgSerLeuArgLeuGlnAlaGlnAsnProSerProAlaPro 159
 357 AGAAAGAAAGAGAGAGACAGATTCAGGGGAAAG..... 395
 159 erProSerArgTrgAspAlaAsn.ProSerAspGluAspAspGlyVa 175
 396GGGAGAAACGG...AGAGAGTTAAGAG..... 422
 175 lGlySerGlyArgArgArgGlnLysArgLeuLysSerValLeuSer 192
 422 422
 192 erSerGlyGlyGluArgPheAsnProPheHisArgArgGluGlnTyr 208
 422 422
 209 AlaAspLeuThrSerPheAspGlyPheAspLeuPheProGlyLeuLe 225
 423GATMAAAGACGAGATCGAGAC..... 446
 225 uCysCysLysArgCysProGlyLeuLysLysArgGlnLysSerSerPhe 242
 446 446
 242 heValCysGlnValTrpSerPheAlaAlaLeuArgProGlySerArgLys 258
 447CGGCTGAGATGAGCGAGAAAGATCTCCAGTG 481
 275 sSerMetLeuGlnAspAspGluAlaProAlaAlaValValLysVal 292
 259 ArgLeuThrPheLeuArgPheHisGlyGlnPheAspSerAspGlyCys 275
 482 T.....CAGGCCCT.....GTGAGAT 498
 499 TAGACTGCTCCTCAGAAAGCTTCACAAAGCTTTAGCCAAACAGAA 548
 292 alGluValGlnGluLysLysLysValSerSerLysAlaThrGlyLys 308
 549GAGTAGAAGACAGACCCCT..... 569

DR PROSITE: P550014: BROMODOMAIN_2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 556 AA; 62060 MW; 03878B1E71C891C2 CRC64;

alignment_scores:
 Quality: 236.00 Length: 447
 Ratio: 1.078 Gaps: 20
 Percent Similarity: 48.993 Percent Identity: 24.161

alignment_block:
 US-09-687-230-1 x Q9SFX1 ..

Align seg 1/1 to: Q9SFX1 from: 1 to: 556

```

172 AGCACAGAAAGCAGACAGTGGGAGCAAAACACCTTACGAGAGATATGTAGAG 221
   |||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
51 SerSerArgAsnProAsnPhSerAsnArgSerAsnArgArgArgArgArgArgArg 67
   ::|||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
222 AAGCCCTTGAGCTGGTCTCAAAAGTAGAGGAGGAGAGAGAGAGAGAGAGAGAG 271
   ::|||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
67 eArgnSerGlnAspAspAspAspAspGlnArgArgArgArgArgArgArgArgArg 83
   |||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
272 CTCACAGGGCAGCTGGGGCAGCAGCTCCAGCCTCTTCGAAAGCAAAAGAG 321
   |||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
84 LeuHis.GlyLeuAsnSerHisGly..... 92
322 ATCATGACAAACACAGACAGACAGACAGACAGACAGACAGACAGACAG 371
   |||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
93 .....ArgAspSerSerAsnSerLysSerGlyGlyLysArg 104
372 AAGCAGATTCCAGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 421
   ::|||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
105 Leu.....AspSerAspAlaArgAsnArgArgArgArgArgArgArgArg 115
422 GGATAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 471
   |||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
116 .....AspGlySerAspAsnThrGlyGlyLys 124
472 AATCTCAGTGTACGCCCTGTGAGATTAGACTTCCTCTCGAGAGAGCT 521
125 .....AlaSerLysAla 128
522 CTCACAGACTCTTATAGCCAAACAGAGAGAGAGAGAGAGAGAGAGAG 569
   |||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
129 ThrAspIleLeuGlnArgGlySerLeuValGlnSerThrProLeuPr 145
570 .....CAGAGAGCTTGATCAACAGAGAGAGAGAGAGAGAGAGAGAG 615
   ::|||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
145 AspLysLysLeuLeuPhePheIleLeuAspArgValGlnLysLysAsp 162
616 CAGCTGCTTCTTTCATTTCTCTGACTGATTTATTTGCTCTGGCTAC 665
   ::|||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
162 hrTyGlyValLysSerAspProAlaAspProGlnLeuLeuProAspTyr 178
666 TCATGATCATTAACACCCATGATTTAGTACATGAAAGAGAGAGAGAT 715
   |||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
179 TyrGlnIleIleLysAsnProMetAspPheThrLeuArgLysLys 195
716 CAGAGACATGACTATCATGCTCATGAGAGAGAGAGAGAGAGAGAGAG 753
   ::|||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
195 uGlnSerGlyValLysThrThrLeuGlnPheGlnAlaSerLeuGlnA 212
754 ATTAACCTTCAAACTAATGTACTAATGCCATGATTTACATTAACACAG 803
   |||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
212 spValPhe...LeuIleCysThrAsnAlaMetGlnTyrAsnSerAlaAsp 227
804 ACATTTATATATAAGCTGCAAGAGAGAGCTTGACACTCAGAGAGAGAGAT 853
   |||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
228 ThrValTyrTyrArgGlnAlaArgAlaMetLeuGlnLeuAlaLysLys 243
854 TCTTAGCCAGAGAGAGAGATTCAGAGCTGAGAGAGAGAGAGAGAGAGAT 903

```

```

243 ..... 243
904 CTGACTTGACAGAAACTCGAAAGCAGAAAGATGCAACAGACCTTCACAG 953
   |||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
244 ..AspPheGlyAsnLeuArgGlnGlnSerAspGlyGlnGlnProValSer 259
954 AGTGGGAGAGAGAGAGCTGCTGGCAGAGAGAGAGAGAGAGAGAGAGAG 1003
   ::|||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
260 LeuSerGlnGlnProLysValValLysArgGlyArg..... 271
1004 TGCCGAGACACAGCCTTCAAGAGTCCAGAGAGAGAGAGAGAGAGAGAG 1053
   |||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
272 .....ProProGlnSerLysLeuLysLys 279
1054 ACAAGATATGCTTGAAGATTAAGTTTAAACATTAATTGACAGAGAG 1103
   ::|||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
280 .....GlnLeuGln 282
1104 CAGAGAGAGCTTGAACCGCATGTGAAGAGATCTGAGAGAGAGCTG..... 1148
   |||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
283 GlnSerLeuIleAspArgThrThrSerAspIleSerAlaAspAlaAla 1299
1149 .....ACCAGCGGCTTGTGAACAGTCAAGTGCAGAT 1179
299 aPheThrTyrAlaGlyAspSerSerArgLeuSerGlySer.....T 313
1180 TTGAAGAGAGAGAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAG 1229
   ::|||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
313 TyrAsnLeuArgLys.....AsnProSerTyrGlyPheArgHisAla 327
1230 GTGATCCCATTTGAGAGAGCCAGGCTACTGCTGGTG..... 1268
328 GlnThrSerAlaValAsnLysTyrGlyMetLysAsnValAspGlnAsn 344
1269 .....AGACTGGGATGACAACTGCAAGACTTCAG..TCTGAGAGTGA 1308
344 gArgAspThrTyrAsnGlnAsnSerAlaSerLeuGlnAspSerLys 361
1309 ATACTTTGACAGGGGTTCAAGAGAGATTAAGAGAGAGAGAGAGAGAG 1358
   |||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
361 heThrLeu.....LeuAspAspAsnLeuLysGlnLeuThrProVal 374
1359 ..TTATATTGAATTTATGGCCCTACAGTTCTTATGACACCGCATATGA 1405
   |||::|||::: |||::|||::: |||::|||::: |||::|||::: |||::|||:::
375 GlyLeuLysAlaGlnTyrGlyTyrAlaArgSerLeuAlaArgTyrAlaAl 391
1406 CTCC.....ACATTTGCAATATC 1424
391 aAsnIleGlyProValAlaThrThrPheAlaAsnVal 403

```

THIS PAGE BLANK (USPTO)